

## **Durham Research Online**

---

### **Deposited in DRO:**

22 July 2020

### **Version of attached file:**

Published Version

### **Peer-review status of attached file:**

Peer-reviewed

### **Citation for published item:**

Wan, Ying-Wooi and Al-Ouran, Rami and Mangleburg, Carl G. and Perumal, Thanneer M. and Lee, Tom V. and Allison, Katherine and Swarup, Vivek and Funk, Cory C. and Gaiteri, Chris and Allen, Mariet and Wang, Minghui and Neuner, Sarah M. and Kaczorowski, Catherine C. and Philip, Vivek M. and Howell, Gareth R. and Martini-Stoica, Heidi and Zheng, Hui and Mei, Hongkang and Zhong, Xiaoyan and Kim, Jungwoo Wren and Dawson, Valina L. and Dawson, Ted M. and Pao, Ping-Chieh and Tsai, Li-Huei and Haure-Mirande, Jean-Vianney and Ehrlich, Michelle E. and Chakrabarty, Paramita and Levites, Yona and Wang, Xue and Dammer, Eric B. and Srivastava, Gyan and Mukherjee, Sumit and Sieberts, Solveig K. and Omberg, Larsson and Dang, Kristen D. and Eddy, James A. and Snyder, Phil and Chae, Yooree and Amberkar, Sandeep and Wei, Wenbin and Hide, Winston and Preuss, Christoph and Ergun, Ayla and Ebert, Phillip J. and Airey, David C. and Mostafavi, Sara and Yu, Lei and Klein, Hans-Ulrich and Carter, Gregory W. and Collier, David A. and Golde, Todd E. and Levey, Allan I. and Bennett, David A. and Estrada, Karol and Townsend, T. Matthew and Zhang, Bin and Schadt, Eric and De Jager, Philip L. and Price, Nathan D. and Ertekin-Taner, Nilüfer and Liu, Zhandong and Shulman, Joshua M. and Mangravite, Lara M. and Logsdon, Benjamin A. (2020) 'Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models.', *Cell reports.*, 32 (2). p. 107908.

### **Further information on publisher's website:**

<https://doi.org/10.1016/j.celrep.2020.107908>

### **Publisher's copyright statement:**

© 2020 The Authors. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### **Additional information:**

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

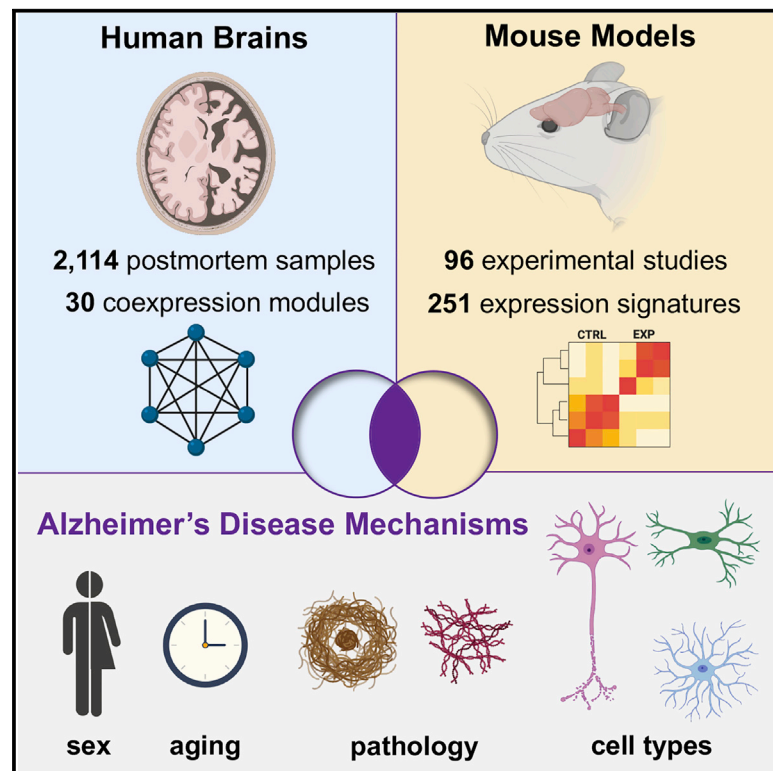
---

Durham University Library, Stockton Road, Durham DH1 3LY, United Kingdom  
Tel : +44 (0)191 334 3042 | Fax : +44 (0)191 334 2971  
<https://dro.dur.ac.uk>

# Cell Reports

## Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models

### Graphical Abstract



### Authors

Ying-Wooi Wan, Rami Al-Ouran, Carl G. Mangleburg, ..., Joshua M. Shulman, Lara M. Mangravite, Benjamin A. Logsdon

### Correspondence

zhandonl@bcm.edu (Z.L.),  
joshua.shulman@bcm.edu (J.M.S.),  
lara.mangravite@sagebionetworks.org (L.M.M.),  
ben.logsdon@sagebionetworks.org (B.A.L.)

### In Brief

Wan et al. develop a consensus atlas of the human brain transcriptome in Alzheimer's disease, based on 2,114 postmortem samples, and examine overlap with 251 expression signatures from mouse brain RNA sequencing experiments, including many neurodegenerative disease models.

### Highlights

- Analysis of 2,114 postmortem samples reveals the Alzheimer's brain transcriptome
- Overlap with mouse models pinpoints age-, sex-, and pathology-induced changes
- Cross-species resource highlights conserved transcriptional networks in disease



## Resource

# Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models

Ying-Wooi Wan,<sup>1,2,43</sup> Rami Al-Ouran,<sup>2,3,43</sup> Carl G. Mangleburg,<sup>1,2,43</sup> Thanneer M. Perumal,<sup>4,43</sup> Tom V. Lee,<sup>2,5</sup> Katherine Allison,<sup>2,5</sup> Vivek Swarup,<sup>6</sup> Cory C. Funk,<sup>7</sup> Chris Gaiteri,<sup>8</sup> Mariet Allen,<sup>9</sup> Minghui Wang,<sup>10</sup> Sarah M. Neuner,<sup>11</sup> Catherine C. Kaczorowski,<sup>11</sup> Vivek M. Philip,<sup>11</sup> Gareth R. Howell,<sup>11</sup> Heidi Martini-Stoica,<sup>12</sup> Hui Zheng,<sup>1,12</sup> Hongkang Mei,<sup>13</sup> Xiaoyan Zhong,<sup>13</sup> Jungwoo Wren Kim,<sup>14,15,42</sup> Valina L. Dawson,<sup>14,15,16,17,19</sup> Ted M. Dawson,<sup>14,16,17,18,19</sup> Ping-Chieh Pao,<sup>20,21</sup> Li-Huei Tsai,<sup>20,21</sup> Jean-Vianney Haure-Mirande,<sup>22</sup> Michelle E. Ehrlich,<sup>10,22</sup> Paramita Chakrabarty,<sup>23</sup> Yona Levites,<sup>23</sup> Xue Wang,<sup>9,24</sup> Eric B. Dammer,<sup>25</sup> Gyan Srivastava,<sup>26</sup> Sumit Mukherjee,<sup>4</sup> Solveig K. Sieberts,<sup>4</sup> Larsson Omberg,<sup>4</sup> Kristen D. Dang,<sup>4</sup> James A. Eddy,<sup>4</sup> Phil Snyder,<sup>4</sup> Yooree Chae,<sup>4</sup> Sandeep Amberkar,<sup>27,28</sup>

(Author list continued on next page)

<sup>1</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

<sup>2</sup>Jan and Dan Duncan Neurologic Research Institute, Texas Children's Hospital, Houston, TX 77030, USA

<sup>3</sup>Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA

<sup>4</sup>Sage Bionetworks, Seattle, WA 98121, USA

<sup>5</sup>Department of Neurology, Baylor College of Medicine, Houston, TX 77030, USA

<sup>6</sup>Department of Neurobiology and Behavior, University of California, Irvine, CA 92697, USA

<sup>7</sup>Institute for Systems Biology, Seattle, WA 98109, USA

<sup>8</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

<sup>9</sup>Mayo Clinic, Department of Neuroscience, Jacksonville, FL 32224, USA

<sup>10</sup>Department of Genetics and Genomic Sciences, Mount Sinai Center for Transformative Disease Modeling, Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

<sup>11</sup>The Jackson Laboratory, Bar Harbor, ME 04609, USA

<sup>12</sup>Huffington Center on Aging, Baylor College of Medicine, Houston, TX 77030, USA

(Affiliations continued on next page)

## SUMMARY

We present a consensus atlas of the human brain transcriptome in Alzheimer's disease (AD), based on meta-analysis of differential gene expression in 2,114 postmortem samples. We discover 30 brain coexpression modules from seven regions as the major source of AD transcriptional perturbations. We next examine overlap with 251 brain differentially expressed gene sets from mouse models of AD and other neurodegenerative disorders. Human-mouse overlaps highlight responses to amyloid versus tau pathology and reveal age- and sex-dependent expression signatures for disease progression. Human coexpression modules enriched for neuronal and/or microglial genes broadly overlap with mouse models of AD, Huntington's disease, amyotrophic lateral sclerosis, and aging. Other human coexpression modules, including those implicated in proteostasis, are not activated in AD models but rather following other, unexpected genetic manipulations. Our results comprise a cross-species resource, highlighting transcriptional networks altered by human brain pathophysiology and identifying correspondences with mouse models for AD preclinical studies.

## INTRODUCTION

Alzheimer's disease (AD) is a progressive and incurable neurodegenerative disorder, with rapidly increasing prevalence due to population aging and no disease-modifying interventions (Scheltens et al., 2016). At autopsy, AD is characterized by extracellular amyloid plaques and intraneuronal neurofibrillary tan-

gles, composed of misfolded and aggregated amyloid-beta (A $\beta$ ) peptide and microtubule-associated protein tau (MAPT or tau), respectively. The mechanisms by which this AD pathologic cascade is triggered and propagated remain incompletely defined and are likely influenced by heterogeneous and dynamic perturbations involving multiple biological pathways. Besides synaptic and neuronal loss, AD pathology is also accompanied



Wenbin Wei,<sup>27,29</sup> Winston Hide,<sup>27,30</sup> Christoph Preuss,<sup>11</sup> Ayla Ergun,<sup>31</sup> Phillip J. Ebert,<sup>32</sup> David C. Airey,<sup>32</sup> Sara Mostafavi,<sup>33</sup> Lei Yu,<sup>8</sup> Hans-Ulrich Klein,<sup>34,35</sup> Accelerating Medicines Partnership-Alzheimer's Disease Consortium<sup>36</sup>, Gregory W. Carter,<sup>11</sup> David A. Collier,<sup>37</sup> Todd E. Golde,<sup>23</sup> Allan I. Levey,<sup>38</sup> David A. Bennett,<sup>8</sup> Karol Estrada,<sup>31</sup> T. Matthew Townsend,<sup>39</sup> Bin Zhang,<sup>10</sup> Eric Schadt,<sup>10</sup> Philip L. De Jager,<sup>34,35</sup> Nathan D. Price,<sup>7</sup> Nilüfer Ertekin-Taner,<sup>9,40</sup> Zhandong Liu,<sup>2,3,44,45,\*</sup> Joshua M. Shulman,<sup>1,2,5,12,41,44,45,\*</sup> Lara M. Mangravite,<sup>4,44,45,\*</sup> and Benjamin A. Logsdon<sup>4,44,45,46,\*</sup>

<sup>13</sup>Neuroscience DPU, Shanghai R&D, GlaxoSmithKline, Shanghai, China

<sup>14</sup>Neuroregeneration and Stem Cell Programs, Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>15</sup>Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>16</sup>Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>17</sup>Adrienne Helis Malvin & Diana Helis Henry Medical Research Foundations, New Orleans, LA 70130, USA

<sup>18</sup>Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>19</sup>Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>20</sup>The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>21</sup>Broad Institute of Harvard University and the Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>22</sup>Departments of Neurology and Pediatrics, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

<sup>23</sup>Evelyn F. and William L. McKnight Brain Institute, Center for Translational Research in Neurodegenerative Disease, Department of Neuroscience, University of Florida, Gainesville, FL 32610, USA

<sup>24</sup>Mayo Clinic, Department of Health Sciences Research, Jacksonville, FL 32224, USA

<sup>25</sup>Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

<sup>26</sup>Data & Statistical Sciences, AbbVie, Cambridge, MA, USA

<sup>27</sup>Sheffield Institute of Translational Neuroscience, University of Sheffield, Sheffield, S10 2HQ, UK

<sup>28</sup>Molecular Oncology Lab, Cancer Research UK – Manchester Institute, The University of Manchester, Manchester, SK10 4TG, UK

<sup>29</sup>Department of Biosciences, Durham University, Durham, DH1 3LE, UK

<sup>30</sup>Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA

<sup>31</sup>Translational Genome Sciences, Biogen, Cambridge, MA, USA

<sup>32</sup>Eli Lilly & Company, Lilly Corporate Center, Indianapolis, IN 46285, USA

<sup>33</sup>University of British Columbia, Vancouver, BC, Canada

<sup>34</sup>Center for Translational & Computational Neuroimmunology, Department of Neurology and Taub Institute for the Study of Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, NY 10032, USA

<sup>35</sup>Cell Circuits Program, Broad Institute, Cambridge, MA 02142, USA

<sup>36</sup>Full consortium author list available at <http://doi.org/10.7303/syn17114455>

<sup>37</sup>Eli Lilly & Company, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey, GU20 6PH, UK

<sup>38</sup>Department of Neurology, Emory University School of Medicine, Atlanta, GA 30322, USA

<sup>39</sup>Foundation Neuroscience Center, AbbVie, Cambridge, MA, USA

<sup>40</sup>Mayo Clinic, Department of Neurology, Jacksonville, FL 32224, USA

<sup>41</sup>Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA

<sup>42</sup>Present address: Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>43</sup>These authors contributed equally

<sup>44</sup>These authors contributed equally

<sup>45</sup>Senior author

<sup>46</sup>Lead Contact

\*Correspondence: [zhandonl@bcm.edu](mailto:zhandonl@bcm.edu) (Z.L.), [joshua.shulman@bcm.edu](mailto:joshua.shulman@bcm.edu) (J.M.S.), [lara.mangravite@sagebionetworks.org](mailto:lara.mangravite@sagebionetworks.org) (L.M.M.), [ben.logsdon@sagebionetworks.org](mailto:ben.logsdon@sagebionetworks.org) (B.A.L.)

<https://doi.org/10.1016/j.celrep.2020.107908>

by changes in astrocytic and microglial cells, and human genome-wide association studies (GWASs) support a causal role for immune mechanisms (Jansen et al., 2019; Kunkle et al., 2019; Raj et al., 2014). Results from RNA sequencing (RNA-seq) of human postmortem brain tissue further reinforce the importance of microglial and inflammatory mechanisms in AD pathogenesis, among other pathways (Allen et al., 2018; Conway et al., 2018; McKenzie et al., 2017; Patrick et al., 2017; Wang et al., 2016, 2019; Zhang et al., 2013).

Compared with transcriptome profiles in human brains, mouse models allow controlled experimental manipulations that can be used to establish causation, isolate the effects of specific molecular lesions, and investigate for dynamic, age-dependent changes. A large number of AD mouse models have been exten-

sively characterized, and these systems have contributed to our understanding of disease pathogenesis (Götz et al., 2018; Janikowsky and Zheng, 2017; LaFerla and Green, 2012). The most widely used transgenic models express mutant forms of the amyloid precursor protein (APP) gene, with or without presenilin-1/2 (PSEN1/2), which are associated with autosomal-dominant, early-onset AD, or, alternatively, *MAPT* mutations, which cause familial frontotemporal dementia (FTD). These models recapitulate features of AD neuropathology, including plaques or tangles, along with variable degree of neuronal dysfunction or loss and progressive behavioral impairment (Ballatore et al., 2007; Esquerda-Canals et al., 2017). More recently, gene expression profiling, including RNA-seq, has been applied to elucidate brain transcriptome signatures in mouse models. These investigations

**Table 1. AD Cases and Controls Included in This Study**

Study	Total Subjects	Tissue	AD		Control		Total Samples
			Female	Male	Female	Male	
Mayo	179	CBE	47	32	35	37	151
		TCX	49	31	35	36	151
MSSM	164	FP	63	27	23	22	135
		IFG	55	24	17	20	116
		PHG	47	18	18	20	103
		STG	57	28	20	17	122
ROSMAP	241	DLPFC	109	46	47	39	241
Total	584		318	160	148	152	778

Counts (n) of AD case and control subjects and derived samples used for differential expression analysis. CBE, cerebellum; TCX, temporal cortex; FP, frontal pole; IFG, inferior frontal gyrus; PHG, parahippocampal gyrus; STG, superior temporal gyurs; DLPFC, dorsolateral prefrontal cortex. See [Table S1](#) for details of the 2,114 samples included in the coexpression network analyses.

have similarly highlighted the potential importance of immune or inflammatory and neuronal or synaptic changes ([Boisvert et al., 2018](#); [Castanho et al., 2020](#); [Cummings et al., 2015](#); [Gjoneska et al., 2015](#); [Matarin et al., 2015](#); [Rothman et al., 2018](#); [Stephenson et al., 2018](#); [Swartzlander et al., 2018](#)). Although some studies have identified selected overlaps in expression changes between AD mouse models and human brains ([Bennett et al., 2018](#); [Castillo et al., 2017](#); [Mostafavi et al., 2018](#); [Neuner et al., 2019](#); [Rojo et al., 2017](#)), other studies have questioned the overall degree of conservation ([Burns et al., 2015](#); [Galatro et al., 2017](#); [Hargis and Blalock, 2017](#)).

The Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) Consortium has generated RNA-seq profiles from more than 1,200 human brains and is applying systems biology approaches toward the goal of elucidating AD mechanisms and potential therapeutic targets. Here, we perform a meta-analysis including all available AMP-AD RNA-seq datasets and systematically define correspondences between gene expression changes associated with AD in human brains and those caused by controlled experimental manipulations in mouse models, including studies relevant to AD, other neurologic disorders, brain health, and aging. Our results constitute a powerful reference of transcriptome perturbations in AD, highlight sex as an important driver, and identify strengths and limitations among currently available mouse models.

## RESULTS

### A Consensus Atlas of AD-Associated Transcriptional Modules from Human Brain

Because transcript abundance is highly influenced by multiple sources of technical and biological variation, we first evaluated the consistency and robustness of findings across three independent human brain transcriptome studies (ROSMAP [Religious Orders Study and the Memory and Aging Project], MSSM [Mount Sinai School of Medicine], and Mayo), consisting in total of 2,114 samples from 1,234 subjects ([Table S1](#)). Human postmortem brain RNA-seq data were obtained from seven distinct regions: dorsolateral prefrontal cortex (DLPFC), temporal cortex (TCX), inferior frontal gyrus (IFG), superior temporal gy-

rus (STG), frontal pole (FP), parahippocampal gyrus (PHG), and cerebellum (CBE). We applied a stringent definition of AD (n = 478 cases), requiring both a clinical diagnosis and neuropathologic confirmation ([Table 1](#)). All control brains (n = 300) were devoid of significant AD pathology, and in two of the three cohorts, control subjects were clinically confirmed as non-demented. To identify transcriptional modules associated with AD, we integrated differential expression and network coexpression analyses. First, we performed both random- and fixed-effects meta-analyses ([DerSimonian and Laird, 1986](#); [Mantel and Haenszel, 1959](#)) of differential expression across the brain regions, defining as many as 2,355 upregulated and 2,130 down-regulated transcripts in AD brains ([Table 2](#); [Supplemental Information](#)). Next, we performed coexpression analysis using five distinct algorithms (see [STAR Methods](#)), generating 2,978 brain region-specific coexpression modules (CBE, n = 498; DLPFC, n = 450; FP, n = 393; IFG, n = 429; PHG, n = 370; STG, n = 336; TCX, n = 502) ([Supplemental Information](#)). As expected, similar overall coexpression was observed in each brain region dataset independent of the algorithm used, based on gene overlap among modules ( $p_{adj} < 0.05$ , Fisher's exact test). In order to define consensus modules that were differentially expressed in AD, we first restricted our analysis to individual modules that were significantly enriched for the differentially expressed genes (DEGs) from the meta-analysis ( $p_{adj} < 0.05$ , Fisher's exact test from either the fixed- or random-effects model results). We next applied graph clustering ([Pons and Latapy, 2005](#)) to define brain region-specific consensus modules based on patterns of shared module gene membership. This analysis culminated in 30 AD-associated modules across the seven tissue types (CBE, n = 4; DLPFC, n = 4; FP, n = 4; IFG, n = 4; PHG, n = 5; STG, n = 4; TCX, n = 5), with module sizes ranging from 504 to 4,673 genes (mean 2,090) ([Figures S1A–S1E](#)).

Evaluation across tissues demonstrated that these 30 modules fell into five distinct "consensus clusters" that were highly preserved across study and tissue type and demonstrated minimal overlap (henceforth denoted clusters A–E) ([Figure 1A](#); [Figure S1](#)). Indeed, further analysis of the consensus clusters revealed significant conservation across brain regions, with clusters A, B, and C each demonstrating a mean gene overlap



**Table 2. Differential Gene Expression Meta-Analysis Results**

Sex	Meta-Analysis Model	Upregulated Genes	Downregulated Genes
All	fixed	2,355	2,130
	random	1,773	1,577
Females	fixed	3,114	2,847
	random	2,211	2,036
Males	fixed	1,450	1,210
	random	845	567

The total number of AD-associated up- or downregulated genes are noted (false discovery rate [FDR] < 0.05 and mean expression fold change > 0.2) from joint and sex-stratified meta-analysis.

of 83% (SE 0.035) among the constituent modules (i.e., proportion of genes that are shared at least once across modules in the cluster). Consensus clusters D and E had a slightly lower mean overlap of 77% (SE 0.030) in gene set membership among the constituent modules. Compared with the 30 AD-associated modules, other non-associated coexpression modules were less conserved across brain regions (see Figure S2A; Supplemental Information). Modules also included a minority of “private” genes (not seen in other modules within clusters), identifying potential unique regional signatures across the brain. Coexpression modules derived from CBE showed the highest number of private genes, potentially consistent with its distinct cellular architecture (Figure S1).

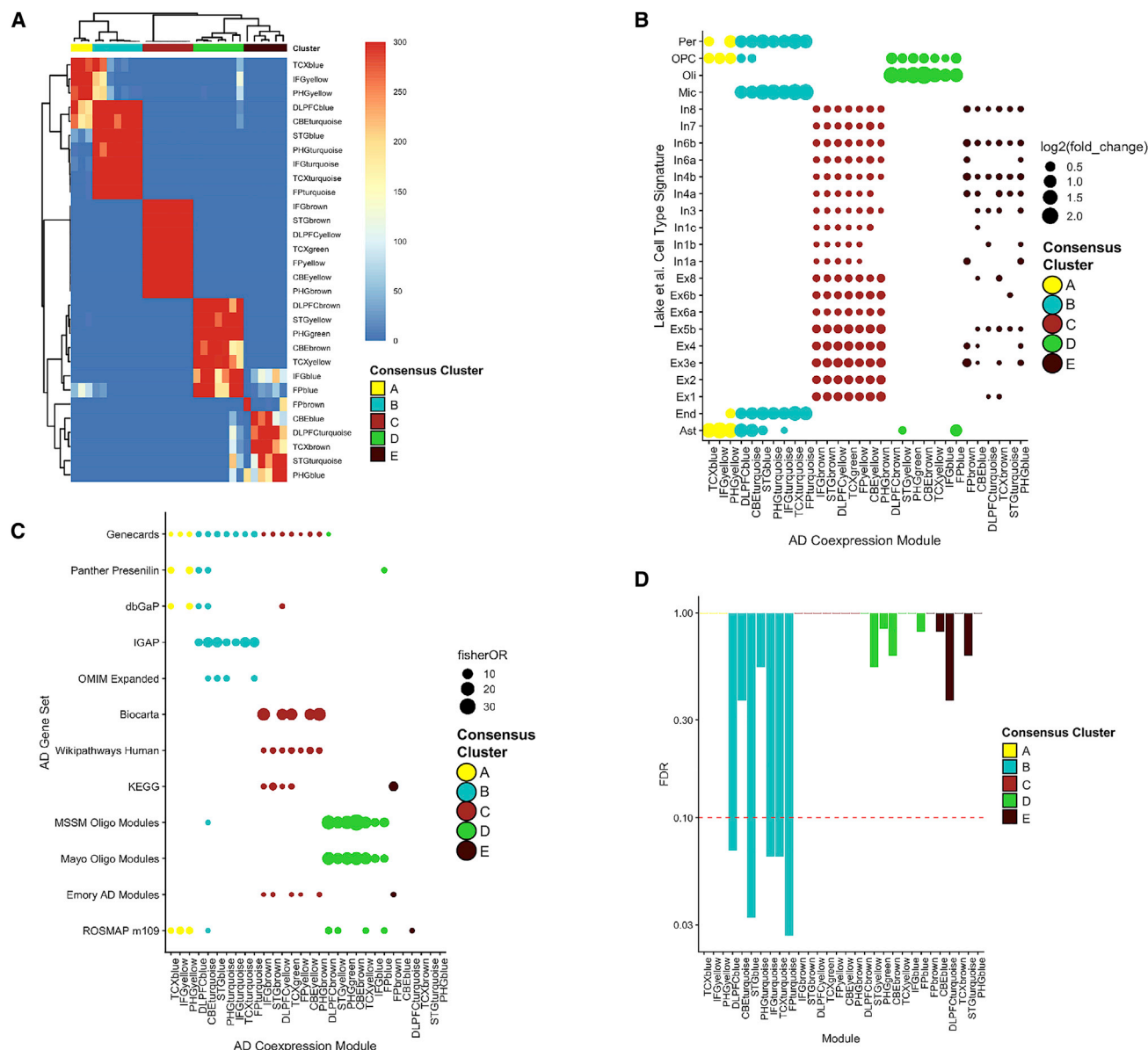
We next evaluated consensus clusters for enrichment of cell type-specific gene signatures (Figure 1B), using expression-weighted cell type enrichment analysis (Skene and Grant, 2016) and a human single-cell RNA-seq reference dataset (Lake et al., 2018). Consensus clusters were enriched for astrocytes and pericytes (clusters A and B), endothelial cells (B), microglia (B), neurons (C and E), oligodendroglia (D), and oligodendroglial precursors (A and D). We obtained consistent results for overlaps using Fisher’s exact test (Figure S2B) and leveraging a complementary cell type reference dataset from mouse brains (Zhang et al., 2014) (Figure S2C). Despite the significant overlaps, we note that cell type-specific signatures constitute a minority of genes within each large cluster. Clusters A–E are also enriched for a diversity of biological processes (Figure 1C; Supplemental Information). Notably, consensus cluster E was enriched for multiple gene sets involved in the regulation of proteostasis, including both the heat shock and unfolded protein responses (Table S2). Last, based on linkage disequilibrium (LD) score regression method in MAGMA (de Leeuw et al., 2015), multiple cluster B modules were significantly enriched for AD GWAS loci (Kunkle et al., 2019) (Figure 1D).

As expected, several of the consensus clusters generated by our meta-analysis recapitulate molecular networks reported in prior analyses restricted to the constituent datasets. For example, independent analyses of RNA-seq from the Mayo and MSSM cohorts previously identified AD-associated coexpression modules enriched for oligodendrocyte expression signatures (Allen et al., 2018; McKenzie et al., 2017), and these gene sets significantly overlapped with consensus cluster D (Figure 1C). We also examined a coexpression module (m109) signif-

icantly associated with both AD pathology and rate of cognitive decline based on an independent analysis of RNA-seq data from the ROSMAP cohort (Mostafavi et al., 2018). Interestingly, m109 showed significant overlap with four of the five consensus clusters (Figure 1C), suggesting that it may represent a coordinated brain transcriptional response involving multiple cell types. Last, we examined an AD-associated coexpression module enriched for RNA-binding proteins that was previously nominated from human brain proteomics (Johnson et al., 2019). This gene set significantly overlapped with cluster C, suggesting that it may be coregulated with neuronal genes.

### Overlaps with AD Mouse Models

To enable cross-species comparisons, we uniformly reprocessed brain RNA-seq data from 96 distinct mouse studies relevant to AD, other neurodegenerative disorders, aging, and related mechanisms (Figure 2; Table S3). We next curated 376 unique experimental comparisons (genetic manipulation versus control condition). After applying quality control filters (gene membership > 10, fold change > 1.2, false discovery rate < 1%; see also STAR Methods), we define 251 sets of significant DEGs comprising brain “gene expression signatures” (mean 1,385 genes, range 10–12,393) characteristic of each mouse model comparison (Table S4; Supplemental Information). These expression profiles encompass 25,181 unique transcripts among 52,873 total in the mouse transcriptome. The curated expression signatures capture a combination of disease-specific and overlapping features characteristic of the heterogeneous mouse models included in this study, on the basis of t-weighted stochastic neighbor embedding (t-SNE) plots (Figure 2C). Even among mouse models within the same disease category, the overall extent of gene overlap among the corresponding sets of DEGs was modest (Figure S3A); therefore, the included comparisons sample a wide spectrum of brain transcriptional responses. We next evaluated the overlap between each mouse expression signature and the 30 human AD-associated brain consensus coexpression modules defined by our meta-analysis. Overall, we detected 1,569 significant overlaps ( $p_{\text{adj}} < 0.01$ ), with the majority (68%) of experimental comparisons having gene expression changes that corresponded to multiple human modules (mean 6 overlapping modules per mouse model DEG set) (Figure 3; Table S5). As expected, most mouse DEG sets showed consistent overlaps with human modules that fell within consensus clusters A–D. We next assessed whether the direction of gene expression changes in mouse and human brains was concordant, that is, whether transcriptional perturbations are phased in the same direction (both up- or downregulated). Indeed, the majority (77%) of overlapping gene expression signatures were concordant in direction across species (Figure 3). For example, cluster B genes were consistently upregulated in human AD brain transcriptomes and mouse AD models, whereas cluster C genes were predominantly downregulated. The exceptional discordant modules (e.g., Fpturquoise and DLPFCyellow in clusters B and C, respectively) may highlight coregulated gene sets with greater temporal or regional fluctuations over the disease course. Last, 28 of 30 human coexpression modules significantly overlapped ( $p_{\text{adj}} < 0.01$ ) with at least one mouse model expression signature. In sum, our results support broad



**Figure 1. Human Consensus RNA-Seq Coexpression Modules**

(A) Gene set overlap (Fisher's exact test p value) was examined among 30 AD-associated coexpression modules, highlighting five consensus clusters (A, B, C, D, and E).

(B) Coexpression modules were evaluated using expression-weighted cell type enrichment analysis based on human single-cell RNA-seq.

(C) Coexpression modules were examined for overlap (Fisher's exact test) with curated AD gene sets from GeneCards, Panther, the Database of Genotypes and Phenotypes (dbGaP), IGAP, Online Mendelian Inheritance in Man (OMIM), Biocarta, Wikipathways, and the Kyoto Encyclopedia of Genes and Genomes (KEGG). We also evaluated overlap with coexpression modules derived from the constituent cohorts, including oligodendroglial modules identified by Mayo and MSSM, module 109 from ROSMAP, and an RNA-binding protein rich module from Emory.

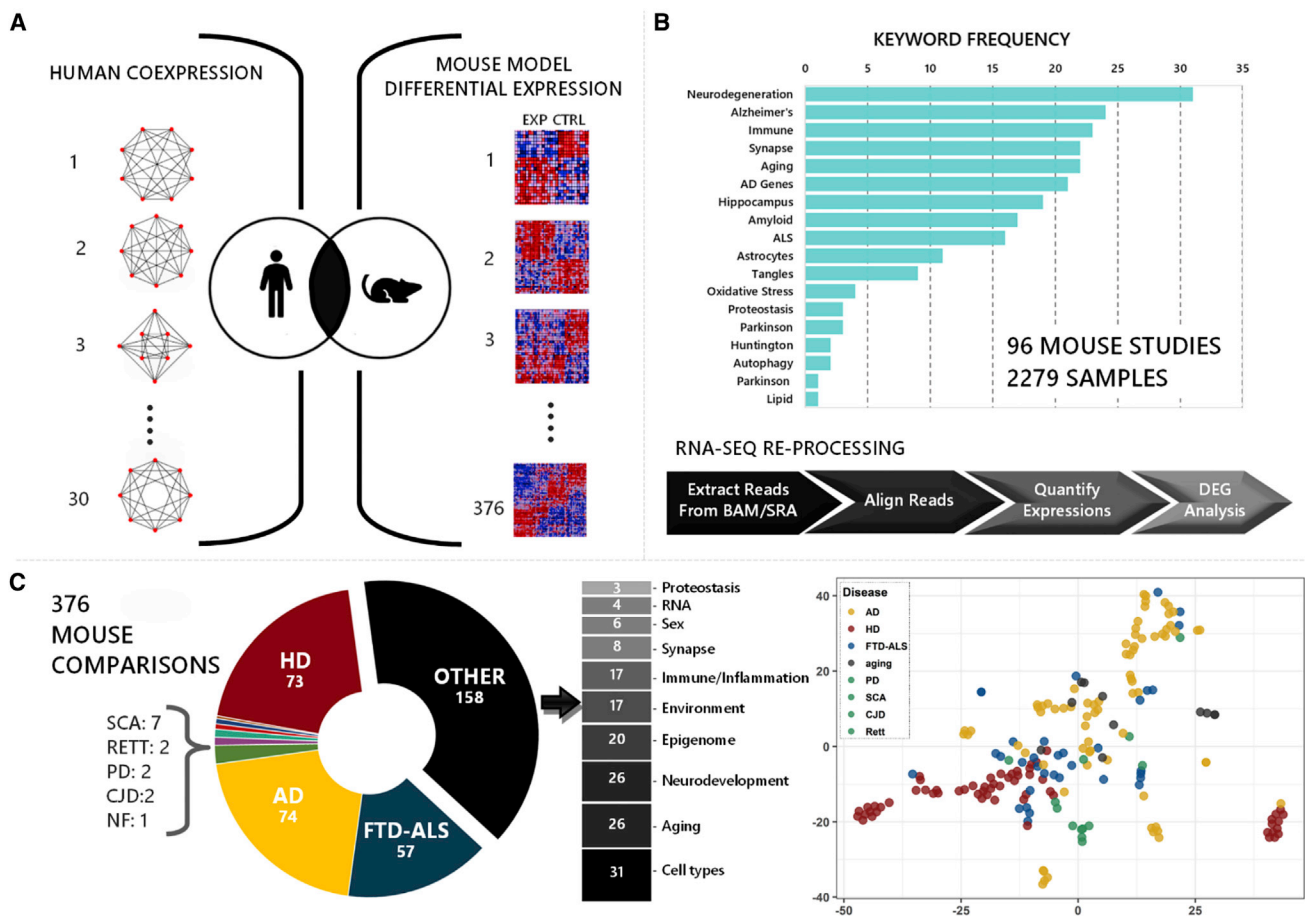
(D) Coexpression module enrichment for AD susceptibility gene candidates from GWAS, based on MAGMA. Consensus cluster B modules appear strongly enriched for AD risk loci.

conservation of gene regulatory systems in the mammalian brain, highlighting that many AD-associated brain expression patterns are recapitulated in mouse experimental models.

Our reprocessed dataset includes 53 expression signatures from 12 distinct AD transgenic models, including multiple *APP* and *MAPT* transgenic strains (Figure 4A). Human microglial-

and neuronal-enriched clusters (B and C, respectively) strongly overlapped with expression signatures from mouse models of AD (Figures 4A–4C). For example, module FPTurquoise from cluster B significantly overlapped ( $p_{adj} < 1 \times 10^{-5}$ ) with the majority of expression signatures from both *APP* (66%) and *MAPT* (67%) mouse models. Although overlaps with cluster C modules





**Figure 2. Cross-Species Study Design and Data**

(A) Analytic design for examining overlaps between 30 Alzheimer's disease (AD)-associated human coexpression modules and differentially expressed gene sets from 376 experimental comparisons in mouse models.

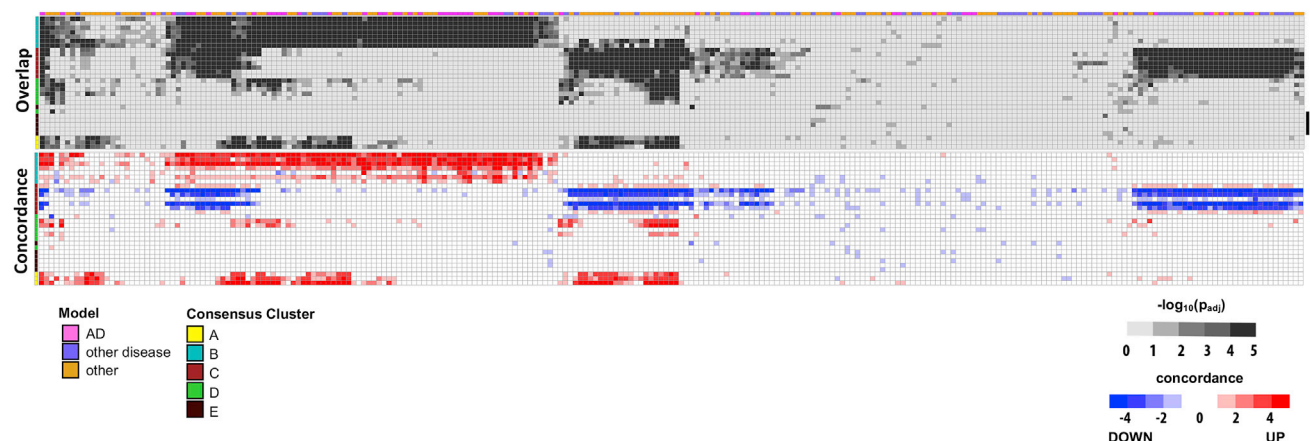
(B) Ninety-six mouse studies were selected based on relevance to AD and other neurodegenerative disorders. Distribution of keywords is shown among all studies. RNA-seq was reprocessed using a standard pipeline. See [Table S3](#) for details on all included mouse studies.

(C) The differentially expressed gene sets represent mouse models of AD, Huntington's disease (HD), frontotemporal dementia-amyotrophic lateral sclerosis (FTD-ALS), spinocerebellar ataxia 1 (SCA1), Rett syndrome (RETT), Parkinson's disease (PD), Creutzfeldt-Jakob disease (CJD), neurofibromatosis (NF), or other neurodegenerative mechanisms. A t-distributed stochastic neighbor embedding (t-SNE) plot including all mouse differential expression signatures highlight both disease-specific and overlapping features among heterogeneous neurodegenerative models.

See also [Figure S3](#).

(e.g., PHGbrown) were more restricted, significant overlaps were still observed in a substantial proportion of AD mouse models (22% and 25% of *APP* and *MAPT* models, respectively). In most cases, attenuation of cluster C gene expression and activation of cluster B appeared to be mutually exclusive ([Figures 4A](#) and [4C](#)). Interestingly, coincident changes in both cluster B and cluster C module genes ([Figure 4A](#), circles) were characteristic of selected *MAPT* strains and were also seen in the *CDK5-P25* mouse, which similarly develops extensive tau pathology and brain atrophy ([Cruz et al., 2003](#)). This overlap pattern was only rarely seen among the *APP* models (1 of 34 DEG sets). Potentially consistent with late microgliosis, *APP* and *MAPT* mouse expression signatures demonstrated stronger overlaps with cluster B modules at more advanced stages of brain pathology ([Figure 4A](#)). Reciprocally, the human cluster C modules over-

lapped expression signatures from mouse models with comparatively mild pathologic burden, suggestive of early-stage synaptic and neuronal dysfunction. RNA-seq profiles generated from multiple aged time points allow examination of dynamic changes in human overlap patterns for selected AD models ([Figure 4C](#)). For example, in the TgCRND8 *APP* transgenic mouse, cluster B module activation was seen by 6 months, and this signature was sustained at 12 months. In contrast, in the rTg4510 *MAPT* mouse model, transient gene expression changes overlapping cluster C human modules either preceded or accompanied the appearance of cluster B expression signatures (M239-M244; [Figure 4A](#), arrowheads, and [Figure 4C](#), bottom). Compared with cluster B and C modules, overlaps between AD mouse model expression signatures and clusters A, D, and/or E were more sparsely detected ([Figure 4A](#)). For



**Figure 3. Overview of Human-Mouse Overlaps and Concordance**

Heatmaps show overlap (top) and concordance (bottom) among 30 human coexpression modules (rows) and 251 sets of differentially expressed genes (DEGs; columns) from mouse model comparisons. The average sample size is 8.4 (range 4–28 total samples). Mouse-human overlap significance, calculated using the hypergeometric test, is represented in grayscale ( $-\log_{10}(p_{adj})$ ). Direction (red/blue) and extent of concordance (intensity) for gene expression changes are also indicated (bottom). The color bar at the top annotates all DEGs based on whether they derive from Alzheimer's disease (AD) models (pink), other neurodegenerative models (purple), or other experimental manipulations potentially relevant to AD mechanisms (orange). The color bar at left denotes cluster membership (A–E). Cluster E modules (brown, asterisk) show sparse overlap with AD mouse model DEGs. See Tables S4, S5, and S6 for details on all experimental comparisons and comprehensive results.

example, TCXyellow and STGyellow—two cluster D modules similarly enriched for oligodendroglial expression signatures including genes with roles in sphingolipid metabolism—showed selective overlap with aged *APP* transgenic models (e.g., TgCRND8 [M247], 5xFAD [M223, M224], and PS2APP [M145]) (Figure 4A, squares; Figure S4).

### Sex-Specific Expression Signatures in Humans and Mice

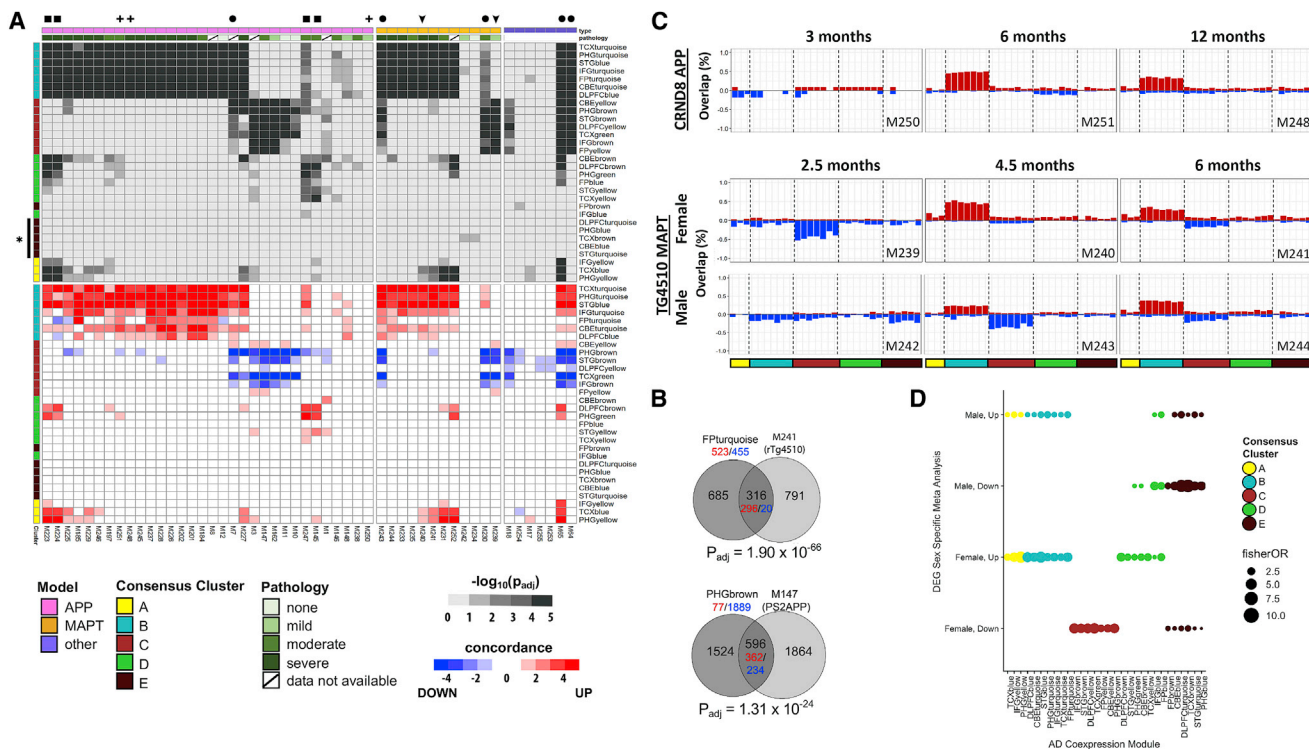
Prior observations in AD mouse models (Jiao et al., 2016) and human epidemiology (Altmann et al., 2014; Li and Singh, 2014; Mayeux and Stern, 2012) have implicated sex as a modifier of AD manifestation and progression. In selected cases, our cross-species analyses provided consistent evidence of a sex-by-age interaction. Compared with males, female rTg4510 mice demonstrated overlaps with human coexpression modules consistent with accelerated progression in brain transcriptional changes (Figure 4C, bottom). Specifically, compared with male mice, for which downregulated genes (M242) do not show overlaps with cluster C modules until 4.5 months, genotype-matched female mice (M239) exhibited similar changes at 2.5 months. Furthermore, beginning at 4.5 months, upregulated genes from females also showed a strong overlap with cluster B modules. These results suggest that sex may modify age- and AD pathology-dependent progression of brain expression changes.

We next repeated our differential expression meta-analysis of human RNA-seq data and stratified by sex (Table 1). Strikingly, we observed  $\sim 2.5$  times more DEGs in women than in men, consistent with other recently published work (Mathys et al., 2019). Sex-specific DEGs also showed evidence of heterogeneous overlaps among the coexpression modules (Figure 4D). For example, clusters C and D were significantly and selectively enriched for female-specific DEGs. Although clusters A and B overlapped with DEGs from either sex, approximately 3-fold increased enrichment was seen for female DEGs compared

with male DEGs. In contrast, male-specific DEGs exhibited preferential overlapping with cluster E. Specifically, upregulated transcripts from male subjects selectively overlapped cluster E modules, whereas downregulated cluster E transcripts showed stronger enrichment for female DEGs. Overall, results from both human brains and mouse models suggest that sex has a strong influence on brain gene expression changes in AD.

### Overlaps with Aging and Other Disease Expression Signatures

Numerous mechanistic parallels have been recognized between AD and other neurodegenerative disorders, including similar protein aggregate pathologies, proteostatic and oxidative stress, neuroinflammation, and the critical role of aging in disease risk and/or progression (Block and Hong, 2005; Bucciantini et al., 2002; Guo and Lee, 2014; Haass and Selkoe, 2007; Nixon, 2005; Ross and Poirier, 2004). We therefore examined for overlap between human AD coexpression modules and expression signatures from mouse experimental models for other neurodegenerative disorders, including Huntington's disease (HD), FTD-amyotrophic lateral sclerosis (ALS), Parkinson's disease, spinocerebellar ataxia 1 (SCA1), Creutzfeldt-Jakob disease (CJD) and Rett syndrome, along with data from aged, wild-type mice mouse strains (Figure 5; Figure S5). Similar to *APP/MAPT* transgenic mice, most other neurodegenerative disease mouse models strongly activated expression signatures overlapping with the cluster B and cluster C modules from human brain. For example, PHGbrown (cluster C) overlaps with mouse brain expression signatures from models of HD (M94,  $p_{adj} = 2.5 \times 10^{-10}$ ), FTD-ALS (*FUS*) (M173,  $p_{adj} = 4.8 \times 10^{-6}$ ), SCA1 (M155,  $p_{adj} = 1.5 \times 10^{-18}$ ), and CJD (M200,  $p_{adj} = 6.0 \times 10^{-32}$ ). Similarly, FPTurquoise (cluster B) significantly overlaps with DEGs from models of FTD-ALS (*TDP43*) (M24,  $p_{adj} = 2.0 \times 10^{-42}$ ), CJD (M199,  $p_{adj} = 3.7 \times 10^{-18}$ ), and Rett syndrome (M198,



**Figure 4. Human Coexpression Module Overlaps with AD Mouse Models**

(A) Heatmaps show overlap (top, hypergeometric test) and concordance (bottom) among human coexpression modules and sets of differentially expressed genes (DEGs) from Alzheimer's disease (AD) mouse models. Mouse-human overlap significance, calculated using the hypergeometric test, is represented in grayscale ( $-\log_{10}(p_{adj})$ ). The color bar at the top denotes APP (pink), MAPT (orange), or other (purple) model comparisons. The estimated pathologic burden (plaques/tangles and neuronal loss) in APP and MAPT models is also annotated in green. Cluster E modules (brown, asterisk) show sparse overlap with AD mouse models. Selected overlaps are denoted as follows: squares, APP models with oligodendrocyte-enriched module overlaps; cross-hatches, CRND8-APP models showing sustained activation of microglial modules from 6 months onward; arrowhead, transient activation of neuronal modules in TG4510-MAPT model preceding microglial module overlap; circles, co-activation of neuronal and microglial modules. See Tables S4 and S5 for details on all experimental comparisons, including sample sizes and genotypes, along with comprehensive results. See Figure S4 for duplicated panel including detailed model annotations. See Figure S6 for additional analysis of overlap specificity.

(B) Representative overlaps of human modules with mouse DEGs. The hypergeometric test was applied to assess significance. Gene counts are noted in black, including for overlapping and non-overlapping regions. To assess concordance between human brains and mouse models, gene counts are shown, noting increased (red) or decreased expression (blue), including for the human coexpression module and the overlapping mouse gene set. M147 was derived from Srinivasan et al. (2016).

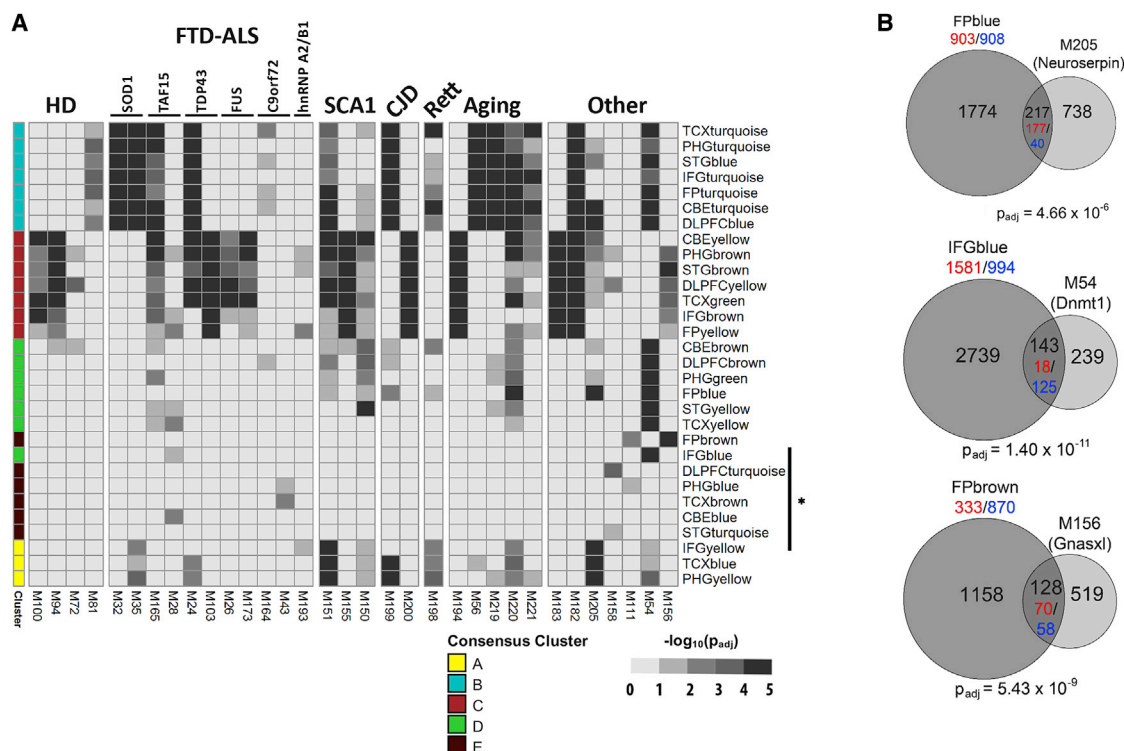
(C) Mouse model overlaps highlight age- and sex-dependent changes. Increasing (red) or decreasing (blue) gene expression and magnitude of changes shown as overlap (%) between the mouse DEG set and module. Cell type module clusters are denoted by colors at panel bottom, as in (A).

(D) Enrichment of human sex-specific DEGs from random-effects meta-analysis among coexpression modules. Consensus cluster C shows downregulation in AD among females.

$p_{adj} = 3.1 \times 10^{-3}$ ) and more selectively in HD (M81,  $p_{adj} = 1.1 \times 10^{-4}$ ) and SCA1 (M151,  $p_{adj} = 5.3 \times 10^{-5}$ ) models. Therefore, these signatures likely represent common brain transcriptome responses induced by diverse neurodegenerative triggers. Indeed, we found that genes implicated in immune biology and inflammation constitute those recurring most frequently among the 251 mouse gene expression signatures included in this study (Figure S3B), consistent with the overlaps seen for cluster B, which is significantly enriched for microglial expression signatures. Importantly, expression signatures from aged, wild-type mice also show significant overlaps with human coexpression modules from clusters B (M56, M219), C (M194), or both B and C concurrently (M220, M221) (Figure 5). For example, compared with 3 month controls, hippocampal tissue from 24-month-old

mice (M56) exhibited expression signatures overlapping with cluster B module FPturquoise ( $p_{adj} = 2.9 \times 10^{-17}$ ). Overall, our results suggest that these patterns may not be specific for neurodegenerative disease, but rather may accompany brain aging more generally. We next examined for any human coexpression modules with overlaps showing relative specificity for AD mouse models. Although none showed absolute specificity, we found that FPblue and TCXyellow, both from the oligodendroglial-enriched cluster D, are strongly activated in selected AD models (e.g., M145, M247, M223, M65 in Figure 4) but show comparatively sparse or weak overlap with expression signatures from the majority of other neurodegenerative disease models. Thus, these modules may encompass transcriptional programs that are preferentially activated by AD pathophysiology.





**Figure 5. Overlaps with Other Mouse Models**

(A) Heatmaps show overlap among human coexpression modules and sets of differentially expressed genes (DEGs) from mouse models, including pure aging, neurodegenerative disorders, and other experimental manipulations. HD, Huntington's disease; FTD-ALS, frontotemporal dementia-amyotrophic lateral sclerosis; SCA1, spinocerebellar ataxia 1; CJD, Creutzfeldt-Jakob disease. Mouse-human overlap significance, calculated using the hypergeometric test, is represented in grayscale ( $-\log_{10}(p_{adj})$ ). Overlaps between HD model expression signatures (Langfelder et al., 2016) and neuronal gene-enriched human coexpression modules recapitulate polyglutamine length (M100, Q92 versus M94, Q175) and brain region dependence (M100/M94, striatum versus M72/M81, cortex). Other manipulations generate signatures similar to AD models, including *PTCH1* knockout (M183) (Ung et al., 2018), *nrf205* (M182) (Ishimura et al., 2016), and *neuroserpin* mutant (M205) (Guadagno et al., 2017). Modules poorly enriched for cell type signatures (asterisk, right) show selected overlaps with FTD-ALS models (M28, M43; Ibrahim et al., 2013, and Lagier-Tourenne et al., 2013, respectively) and other, unexpected genetic manipulations (M158, M111, M54, and M156; Vied et al., 2016, Maze et al., 2015, Narayanan et al., 2014, and Holmes et al., 2016, respectively). See Tables S5 and S6 for comprehensive results, including sample sizes for all comparisons. See Figure S5 for comprehensive heatmaps representing overlaps with HD, FTD-ALS, SCA1, and aging models. (B) Representative overlaps of human modules with mouse DEGs, as in Figure 4B. The hypergeometric test was applied to assess significance.

In order to further assess AD specificity of the consensus modules defined in our analysis, we examined pairwise overlaps with published RNA-seq coexpression modules among 700 human brains from subjects with neuropsychiatric disorders (autism, schizophrenia, bipolar disorder, depression, and alcoholism) (Gandal et al., 2018). Interestingly, virtually all cluster D modules failed to overlap with modules showing significant associations with neuropsychiatric disorders (Figure S6A) and were therefore "AD specific." Similar but less extreme evidence for specificity was seen for cluster E modules. In contrast, AD consensus clusters B and C strongly overlapped with several neuropsychiatric coexpression modules similarly enriched for microglial and neuronal expression signatures, respectively. For example, module CD11, which was increased in autism, overlapped significantly with AD cluster B module FPTurquoise ( $p_{adj} = 1.3 \times 10^{-136}$ ) (Figure S6B). As expected, the genes in CD11 overlapped with DEGs from multiple mouse models (M241,  $p_{adj} = 5.4 \times 10^{-29}$ ; M24,  $p_{adj} = 5.3 \times 10^{-12}$ ; M219,  $p_{adj} = 8.1 \times 10^{-12}$ ), consistent with overlaps seen for consensus cluster B.

Thus, rather than representing specific disease pathophysiology, these modules more likely represent a shared response pattern to diverse forms of brain injury, consistent with their overlaps with expression signatures from heterogeneous mouse disease models and aging. In contrast, other modules show patterns of overlap with mouse AD models that are strikingly divergent from those of the AD consensus clusters. For example, although CD12 strongly overlaps with cluster B, the overlaps with mouse models are more sparse than those seen in CD11. As the modules from Gandal et al. (2018) are much smaller than the AD consensus modules, they may represent subnetworks within the larger consensus clusters that are less strongly activated in AD mouse models.

### Non-obvious Mouse Models of AD

Our analyses also included many additional mouse genetic manipulations relevant to neurodegenerative mechanisms (Figure 2). Indeed, several non-AD mouse models showed overlaps with human AD modules mimicking those of canonical AD mice

(Figure 5). For example, the C57BL6/J (B6J)-*nrf205* mouse (M182) overlaps significantly with modules from consensus clusters B and C (FPturquoise,  $p_{\text{adj}} = 6.4 \times 10^{-6}$ , and PHGbrown,  $p_{\text{adj}} = 9.6 \times 10^{-39}$ , respectively), behaving similarly to *MAPT* transgenics (M230, M243) and *CDK-P25* mice (M64, M65). The C57BL6/J (B6J)-*nrf205* mouse model has mutations in both the translation GTPase *GTPBP2* and the neuronal tRNA<sup>Arg</sup><sub>UCU</sub> genes that cause neurodegeneration through ribosomal stalling (Ishimura et al., 2014, 2016). In another example, expression of a mutant form of *neuroserpin* in mouse neural progenitor cells (M205) that significantly overlapped with cluster A module IF-Gyellow ( $p_{\text{adj}} = 8.1 \times 10^{-17}$ ) as well as FPblue (Figure 5B;  $p_{\text{adj}} = 4.7 \times 10^{-6}$ ), a cluster D module showing some selectivity for AD transgenic models (above). Autosomal-dominant mutations in *neuroserpin* cause familial encephalopathy with Neuroserpin inclusion bodies, a rare, early-onset neurodegenerative dementia caused by protein aggregation within the endoplasmic reticulum (ER) and associated ER stress (Roussel et al., 2013).

Of the 30 human AD-associated coexpression modules, a minority showed virtually no overlap with AD mouse model brain expression signatures (Figure 4A, asterisk). These modules, largely corresponding to consensus cluster E, are strong candidates to represent features of AD pathobiology that are poorly recapitulated by existing AD mouse models. As discussed earlier, these modules showed comparatively poor enrichment for cell-type expression signatures (Figure 1B; Figure S2) and were similarly not well represented among curated AD pathways (Figure 1C), based on gene set enrichment analyses. Interestingly, our cross-species analyses highlight a number of other, unexpected mouse model expression signatures that overlap these modules (Figure 5). For example, conditional knockout of the DNA methyltransferase, *DNMT1* (M54), in the mouse brain (Narayanan et al., 2014) activates a set of DEGs that significantly overlapped ( $p_{\text{adj}} = 1.4 \times 10^{-11}$ ) with a human module, IFGblue, enriched for unfolded protein response and DNA repair pathway genes. In another example, brain RNA-seq from a *Gnasxl*-deficient mouse (Holmes et al., 2016) (M156) overlapped with FPbrown, a coexpression module ( $p_{\text{adj}} = 5.4 \times 10^{-9}$ ) enriched for genes involved in oxidative phosphorylation and mitochondrial translation. Interestingly, *Gnas*, which encodes a G-protein alpha stimulatory subunit, is a complex, imprinted genomic locus implicated in hypothalamic control of energy balance. Loss of the *Gnasxl* isoform causes a hypermetabolic mouse phenotype, resulting in growth retardation, hypoglycemia, and reduced adiposity (Nunn et al., 2013).

## DISCUSSION

We provide a systems-level molecular model of the AD transcriptional state in human brains. Our results highlight five dominant consensus clusters representing robust and reproducible patterns of coexpression patterns in brains affected by AD. These signatures are consistently observed across multiple cohorts and several brain regions and were identified by multiple independent coexpression algorithms. We further define correspondences between 30 human AD brain consensus gene expression

networks and 251 mouse experimental comparisons, including models relevant to AD, other neurologic disorders, and aging. Overall, our meta-analysis of AD-associated gene dysregulation in human brains, along with the complementary cross-species comparisons in mouse models, provides a powerful resource to guide target selection and validation strategies.

Mouse genetic models have contributed enormously to our understanding of AD pathophysiology (Ballatore et al., 2007; Esquerda-Canals et al., 2017); however, the utility of these mice as robust preclinical models for AD has been challenged (Drummond and Wisniewski, 2017; Onos et al., 2016; Sasaguri et al., 2017). First, most AD mouse models are based on rare forms of familial autosomal-dominant AD, which are caused by single, highly penetrant gene mutations. In contrast, late-onset AD arises from dozens of other risk variants, including many with modest effect sizes (Karch et al., 2014; Kunkle et al., 2019), perhaps in combination with non-genetic risk factors. Second, unlike mouse models, most brain autopsies from individuals with AD show evidence of heterogeneous, mixed pathologies that likely modify disease onset, manifestations, and progression (Kapasi et al., 2017). Third, it has been suggested that widely used mouse behavioral assays may be poor predictors of clinically relevant outcomes in humans.

We find that many transgenic mice, including both *APP* and *MAPT* models, manifest gene expression signatures that significantly overlap with AD-associated coexpression modules from human brains. The most robust overlaps were detected among modules enriched for microglial and neuronal genes (clusters B and C). These findings are consistent with prior reports of similar expression signatures detected from human postmortem AD brain tissue (Conway et al., 2018; Grubman et al., 2019; Mathys et al., 2019; Mostafavi et al., 2018; Zhang et al., 2013) or from AD mouse models (Cummings et al., 2015; Gjoneska et al., 2015; Matarin et al., 2015). However, consensus clusters B and C account for only 14 of 30 coexpression modules (47%) and do not appear to be specific for AD pathophysiology. In contrast, a substantial minority of human AD coexpression modules had little to no detectable overlap with available AD mouse models. Among these, consensus cluster E includes signatures for certain inhibitory and excitatory cortical neuronal subtypes and was enriched for genes that regulate proteostasis. The overlaps we define highlight those molecular features of AD biology recapitulated by current mouse models. In contrast, non-overlapping modules may identify dimensions of AD pathophysiology that are poorly captured. We conclude that most AD mouse models show overall poor correspondence to human disease, based on brain transcriptomes, with the exception of neuronal and microglial-enriched coexpression modules. This is an important caveat for the interpretation of studies using these animal models and may explain in part their poor predictive power as preclinical models for AD. Our findings are consistent with a complementary study of mouse data from gene expression arrays (Hargis and Blalock, 2017).

Our analyses also highlight the value of experimental models for interpretation of human brain transcriptome profiles. Analyses considering either cross-sectional or longitudinal datasets similarly suggest that transcriptional changes overlapping human brain neuronal-enriched modules (cluster C) may represent

an earlier, transient stage of AD. In contrast, our results suggest that cluster B modules, strongly enriched for microglial expression signatures, constitute a subsequent and more sustained AD endophenotype. Nevertheless, cluster B genes significantly overlap with candidate susceptibility loci from AD GWASs (Kunkle et al., 2019), which, along with other studies (Sala Frigerio et al., 2019), strongly suggests that although these changes may occur subsequent to disease onset, they are likely causal, perhaps affecting progression. Transition points between human-mouse overlaps can be linked to the manifestation of disease-relevant mouse phenotypes. For example, CRND8 *APP* mice reveal reduced synaptic markers and hippocampal neuronal loss at 6 months, when overlaps are first detected with cluster B modules (Adalbert et al., 2009; Brautigam et al., 2012). Similarly, in Tg4510 *MAPT* transgenics evaluated at 4 and 6 months, respectively, memory task impairment and neurodegenerative pathology correspond to sequential activation of neuronal and microglial expression patterns (Blackmore et al., 2017; Ramsden et al., 2005). As human brain RNA-seq can be evaluated only at the time of death, there are significant challenges to resolve age-dependent changes or to definitively establish links with clinical-pathologic progression. Collection of mouse RNA-seq from additional time points may therefore accelerate discovery of improved AD progression biomarkers and ultimately pinpoint critical windows for therapeutic interventions. Our cross-species approach also highlights the significant impact of sex on the AD brain transcriptome. Based on gene expression profiles, female AD mice progress more rapidly than males, and in a sex-stratified analysis of human brain gene expression, females demonstrated quantitatively greater transcriptional changes. Our results are consistent with prior observations in both mice (Jiao et al., 2016) and humans (Altmann et al., 2014; Li and Singh, 2014; Mayeux and Stern, 2012; Mathys et al., 2019; Sala Frigerio et al., 2019).

Aging is the strongest known AD risk factor. Using a distinct analytic design and largely independent datasets, Hargis and Blalock (2017) reported that DEGs were concordant between aging in humans and rodent models, a conclusion supported by our analysis. Strikingly, the majority of AD-associated human brain coexpression modules overlapping with *APP* and/or *MAPT* transgenic mouse models were also seen in aged, wild-type mice, as well as many other disease models. This result suggests that many human brain gene expression changes associated with AD, including neuronal- and microglial-enriched modules reported in other studies (Conway et al., 2018; Mostafavi et al., 2018; Zhang et al., 2013), may represent common transcriptional programs activated by the aging process itself. Rather than representing a specific signature of AD pathophysiology (e.g., A $\beta$ - or tau-mediated mechanisms), these pathways appear to be activated by heterogeneous triggers, including those manipulated in mouse models of HD, ALS, SCA1, and other neurodegenerative disorders. Interestingly, several module overlap patterns still revealed possible disease-specific signatures. For example, several modules enriched for oligodendroglial markers (consensus cluster D) showed relatively specific overlap with AD mouse models, particularly *APP* transgenic models, and related human brain coexpression networks have previously

been implicated in AD in multiple studies (Allen et al., 2018; McKenzie et al., 2017; Mostafavi et al., 2018). Alternatively, coincident activation of both neuronal- and microglial-enriched modules was seen preferentially in models characterized by significant tau pathologic burden. In contrast to AD and other neurodegenerative disease models, nearly all differential expression signatures from HD mice (35 of 37) (Langfelder et al., 2016) failed to overlap with microglial-enriched coexpression modules (Figure S5).

Our study has several notable limitations. First, as the overall correlation of the brain transcriptome and proteome is modest (Seyfried et al., 2017), it will be important in future work to consider whether human-mouse gene expression overlaps are improved at the protein level. Second, although our meta-analytic approach was designed to identify signatures of disease that are robust to technical and study-specific heterogeneity, it is possible that we could miss relevant gene sets with more modest coexpression or uneven representation across studies because of sample ascertainment differences. Furthermore, the resulting large size of the human consensus modules may limit sensitivity to detect significant overlaps using the hypergeometric test, especially for functional pathways represented by smaller gene sets. Compared with the laboratory preparation of mouse mRNA, the extraction and processing of human brain tissue is likely more susceptible to postmortem artifact, although our analyses adjusted for sample variability in postmortem interval. Moreover, the human RNA-seq data, along with the majority of included mouse studies, derive from bulk brain tissue, which includes mixed cell types. Indeed, most of the human coexpression modules are strongly enriched for cell type-specific signatures, which may therefore reflect global changes in cell proportions, such as neuronal loss or microgliosis. The growing availability of single-cell expression profiles can definitively address this concern. In addition, as the extent and tempo of neurodegeneration, including in both human and mouse models, can vary widely across different brain regions, RNA-seq profiles from whole brain might obscure more localized transcriptome overlaps. Last, although we selected nearly 100 independent mouse RNA-seq studies for inclusion in our analyses, prioritizing those most relevant to AD and neurodegeneration, we omitted many others with the potential to provide additional insights. In the future, our approach can thus be generalized to an even broader sample of available mouse data.

The strengths of our study include coexpression modules based on more than 2,000 human brain samples and consideration of a large and diverse number of mouse studies. All mouse RNA-seq data were reprocessed using a single, complementary pipeline to facilitate cross-species and cross-model comparisons. Importantly, unlike comparisons based on pathology or behavioral phenotypes, brain expression profiles likely represent more proximal endophenotypes, potentially affording greater sensitivity and reliability for detection of cross-species overlaps. In fact, we highlight several overlaps with transcriptomic endophenotypes from completely unexpected mouse experimental manipulations that manifest brain expression changes that mimic human AD and in some cases even overlap coexpression modules better than currently



available AD mouse models. Such “AD transcriptologs”—mouse models based on transcriptome homology—may pinpoint non-obvious experimental models for future investigation of AD pathophysiology.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead Contact
  - Materials Availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Human subject data
  - Mouse Experimental Model Data
- **METHOD DETAILS**
  - Human RNA-Seq Reprocessing, library normalization and covariates adjustment
  - Network Inference and Module Identification
  - Aggregate human coexpression module identification
  - Mouse RNA-seq re-processing
  - Nomenclature for annotation of mouse differentially expressed gene sets
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Meta-Differential expression analysis on human expression profiles
  - Human coexpression module enrichment analysis
  - Differential expression analysis of mouse RNA-seq/microarray
  - Statistical analysis of mouse-human overlaps

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.107908>.

## ACKNOWLEDGMENTS

The results published here are in part based on data obtained from the AMP-AD Knowledge Portal (<https://doi.org/10.7303/syn2580853>). ROSMAP study data were provided by the Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by National Institute on Aging (NIA) grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, and U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute. Mayo RNA-seq study data were provided by the following sources: the Mayo Clinic Alzheimer’s Disease Genetic Studies, led by Dr. Nilufer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, Florida, using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer’s Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, and R01 AG003949; National Institute of Neurological Disorders and Stroke (NINDS) grant R01 NS080820; the CurePSP Foundation; and support from Mayo Foundation. Study data include samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026, National Brain and Tissue Resource for Parkinson’s Disease and Related Disorders), the NIA (P30 AG19610, Arizona Alzheimer’s Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer’s Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901, and 1001 to the Arizona Parkinson’s Disease Consortium), and the Michael J. Fox Foundation for Parkinson’s Research. Mount Sinai Brain Bank (MSBB) data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt of the Mount Sinai School of Medicine through funding from NIA grant U01AG046170. Furthermore, Emory study data were supported through funding from NIA grants P50 AG025688, U01 AG046161, and U01 AG061357. This study was also supported in part by grants from the NIH (R01AG053960, R01AG050631, R01AG057339, U01AG046161, U01 AG046139, RF1AG054014, RF1 AG057440, R01AG057473, R01AG057914, P50 NS38377, and F31AG050357). C.G.M. was supported by the Cullen Foundation and the Baylor College of Medicine Medical Scientist Training Program (MSTP). J.M.S. was additionally supported by grants from the Huffington Foundation, the Jan and Dan Duncan Neurological Research Institute at Texas Children’s Hospital, and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. Z.L. received additional support from the National Science Foundation, Division of Mathematical Sciences (DMS-1263932), the Cancer Prevention Research Institute of Texas (RP170387), the Houston Endowment, and the Neurodegeneration Consortium and Belfer Family Foundation. T.M.D. is supported by the JPB Foundation and is the Leonard and Madlyn Abramson Professor in Neurodegenerative Diseases. The authors acknowledge the joint participation of the Adrienne Helis Malvin Medical Research Foundation through its direct engagement in the continuous active conduct of medical research in conjunction with the Johns Hopkins Hospital and the Johns Hopkins University School of Medicine and the foundation’s Parkinson’s Disease Program (M-2014). We are grateful to Dr. Bruce Yankner for contributing additional mouse RNA-seq data. This study is part of the NIA AMP-AD, MODEL-AD, and Resilience-AD programs.

son’s Disease and Related Disorders), the NIA (P30 AG19610, Arizona Alzheimer’s Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer’s Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901, and 1001 to the Arizona Parkinson’s Disease Consortium), and the Michael J. Fox Foundation for Parkinson’s Research. Mount Sinai Brain Bank (MSBB) data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt of the Mount Sinai School of Medicine through funding from NIA grant U01AG046170. Furthermore, Emory study data were supported through funding from NIA grants P50 AG025688, U01 AG046161, and U01 AG061357. This study was also supported in part by grants from the NIH (R01AG053960, R01AG050631, R01AG057339, U01AG046161, U01 AG046139, RF1AG054014, RF1 AG057440, R01AG057473, R01AG057914, P50 NS38377, and F31AG050357). C.G.M. was supported by the Cullen Foundation and the Baylor College of Medicine Medical Scientist Training Program (MSTP). J.M.S. was additionally supported by grants from the Huffington Foundation, the Jan and Dan Duncan Neurological Research Institute at Texas Children’s Hospital, and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. Z.L. received additional support from the National Science Foundation, Division of Mathematical Sciences (DMS-1263932), the Cancer Prevention Research Institute of Texas (RP170387), the Houston Endowment, and the Neurodegeneration Consortium and Belfer Family Foundation. T.M.D. is supported by the JPB Foundation and is the Leonard and Madlyn Abramson Professor in Neurodegenerative Diseases. The authors acknowledge the joint participation of the Adrienne Helis Malvin Medical Research Foundation through its direct engagement in the continuous active conduct of medical research in conjunction with the Johns Hopkins Hospital and the Johns Hopkins University School of Medicine and the foundation’s Parkinson’s Disease Program (M-2014). We are grateful to Dr. Bruce Yankner for contributing additional mouse RNA-seq data. This study is part of the NIA AMP-AD, MODEL-AD, and Resilience-AD programs.

## AUTHOR CONTRIBUTIONS

Conceptualization, D.A.C., T.E.G., A.I.L., D.A.B., K.E., Z.L., B.Z., E.S., P.L.D.J., N.D.P., N.E.-T., Y.-W.W., J.M.S., B.A.L., T.M.P., C.G., M.W., L.M.M., S.K.S., K.D.D., and P.J.E.; Methodology and Data Curation, Y.-W.W., R.A.-O., T.V.L., K.A., Z.L., B.A.L., T.M.P., V.S., M.W., C.C.F., C.G., M.A., P.S., Y.C., and X.W.; Formal Analysis, Y.-W.W., R.A.-O., C.G.M., B.A.L., J.A.E., K.D.D., P.J.E., and P.S.; Writing – Original Draft, Y.-W.W., C.G.M., R.A.-O., T.V.L., J.M.S., B.A.L., T.M.P., M.A., N.E.-T., L.M.M., A.I.L., D.A.B., P.L.D.J., J.M.S., and G.W.C.; Writing – Review & Editing, C.G.M., R.A.-O., J.M.S., Z.L., T.V.L., K.A., S.M.N., C.C.K., H.Z., V.L.D., A.I.L., B.A.L., T.M.P., M.A., N.E.-T., L.M.M., D.A.B., P.L.D.J., J.M.S., and G.W.C.; Funding Acquisition, J.M.S., Z.L., L.M.M., B.A.L., A.I.L., and T.E.G.; Resources, S.M.N., C.C.K., V.M.P., G.R.H., H.M.-S., H.Z., J.W.K., V.L.D., T.M.D., P.-C.P., L.-H.T., J.-V.H.-M., M.W., M.E.E., H.M., X.Z., P.C., Y.L., T.E.G., B.A.L., and L.M.M.; Supervision, J.M.S., Z.L., B.A.L., and L.M.M.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 2, 2019

Revised: June 1, 2020

Accepted: June 24, 2020

Published: July 14, 2020

## REFERENCES

Adalbert, R., Nogradi, A., Babetto, E., Janeckova, L., Walker, S.A., Kerschene-  
steiner, M., Misgeld, T., and Coleman, M.P. (2009). Severely dystrophic axons  
at amyloid plaques remain continuous and connected to viable cell bodies.  
*Brain* 132, 402–416.

- Adamcsek, B., Palla, G., Farkas, I.J., Derényi, I., and Vicsek, T. (2006). CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22, 1021–1023.
- Ahn, Y.-Y., Bagrow, J.P., and Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature* 466, 761–764.
- Allen, M., Carrasquillo, M.M., Funk, C., Heavner, B.D., Zou, F., Younkin, C.S., Burgess, J.D., Chai, H.-S., Crook, J., Eddy, J.A., et al. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci. Data* 3, 160089.
- Allen, M., Wang, X., Burgess, J.D., Watzlawik, J., Serie, D.J., Younkin, C.S., Nguyen, T., Malphrus, K.G., Lincoln, S., Carrasquillo, M.M., et al. (2018). Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. *Alzheimers Dement.* 14, 352–366.
- Altay, G., and Emmert-Streib, F. (2010). Inferring the conservative causal core of gene regulatory networks. *BMC Syst. Biol.* 4, 132.
- Altmann, A., Tian, L., Henderson, V.W., and Greicius, M.D.; Alzheimer's Disease Neuroimaging Initiative Investigators (2014). Sex modifies the APOE-related risk of developing Alzheimer disease. *Ann. Neurol.* 75, 563–573.
- Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 43, D789–D798.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Ballatore, C., Lee, V.M.-Y., and Trojanowski, J.Q. (2007). Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci.* 8, 663–672.
- Bennett, D.A., Buchman, A.S., Boyle, P.A., Barnes, L.L., Wilson, R.S., and Schneider, J.A. (2018). Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers Dis.* 64 (s7), S161–S189.
- Blackmore, T., Meftah, S., Murray, T.K., Craig, P.J., Blockeel, A., Phillips, K., Eastwood, B., O'Neill, M.J., Marston, H., Ahmed, Z., et al. (2017). Tracking progressive pathological and functional decline in the rTg4510 mouse model of tauopathy. *Alzheimers Res. Ther.* 9, 77.
- Block, M.L., and Hong, J.-S. (2005). Microglia and inflammation-mediated neurodegeneration: multiple triggers with a common mechanism. *Prog. Neurobiol.* 76, 77–98.
- Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008, P10008.
- Boisvert, M.M., Erikson, G.A., Shokhirev, M.N., and Allen, N.J. (2018). The aging astrocyte transcriptome from multiple regions of the mouse brain. *Cell Rep.* 22, 269–285.
- Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H., and Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* 112, 389–404.
- Brautigam, H., Steele, J.W., Westaway, D., Fraser, P.E., St George-Hyslop, P.H., Gandy, S., Hof, P.R., and Dickstein, D.L. (2012). The isotropic fractionator provides evidence for differential loss of hippocampal neurons in two mouse models of Alzheimer's disease. *Mol. Neurodegener.* 7, 58.
- Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., and Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* 416, 507–511.
- Burns, T.C., Li, M.D., Mehta, S., Awad, A.J., and Morgan, A.A. (2015). Mouse models rarely mimic the transcriptome of human neurodegenerative diseases: A systematic bioinformatics-based critique of preclinical models. *Eur. J. Pharmacol.* 759, 101–117.
- Castanho, I., Murray, T.K., Hannon, E., Jeffries, A., Walker, E., Laing, E., Baulf, H., Harvey, J., Bradshaw, L., Randall, A., et al. (2020). Transcriptional signatures of tau and amyloid neuropathology. *Cell Rep.* 30, 2040–2054.e5.
- Castillo, E., Leon, J., Mazzei, G., Abolhassani, N., Haruyama, N., Saito, T., Saido, T., Hokama, M., Iwaki, T., Ohara, T., et al. (2017). Comparative profiling of cortical gene expression in Alzheimer's disease patients and mouse models demonstrates a link between amyloidosis and neuroinflammation. *Sci. Rep.* 7, 17762.
- Clauset, A., Newman, M.E.J., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70, 066111.
- Conway, O.J., Carrasquillo, M.M., Wang, X., Bredenberg, J.M., Reddy, J.S., Strickland, S.L., Younkin, C.S., Burgess, J.D., Allen, M., Lincoln, S.J., et al. (2018). ABI3 and PLCG2 missense variants as risk factors for neurodegenerative diseases in Caucasians and African Americans. *Mol. Neurodegener.* 13, 53.
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.
- Cruz, J.C., Tseng, H.-C., Goldman, J.A., Shih, H., and Tsai, L.-H. (2003). Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. *Neuron* 40, 471–483.
- Cummings, D.M., Liu, W., Portelius, E., Bayram, S., Yasvoina, M., Ho, S.-H., Smits, H., Ali, S.S., Steinberg, R., Pegasiou, C.-M., et al. (2015). First effects of rising amyloid- $\beta$  in transgenic mouse brain: synaptic transmission and gene expression. *Brain* 138, 1992–2004.
- Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23, 1846–1847.
- De Jager, P.L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B.N., Felsky, D., Klein, H.-U., White, C.C., Peters, M.A., Lodgson, B., et al. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* 5, 180142.
- de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 11, e1004219.
- DerSimonian, R., and Laird, N. (1986). Meta-analysis in clinical trials. *Control. Clin. Trials* 7, 177–188.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Drummond, E., and Wisniewski, T. (2017). Alzheimer's disease: experimental models and reality. *Acta Neuropathol.* 133, 155–175.
- Esquerda-Canals, G., Montoliu-Gaya, L., Güell-Bosch, J., and Villegas, S. (2017). Mouse models of Alzheimer's disease. *J. Alzheimers Dis.* 57, 1171–1183.
- Gaiteri, C., Chen, M., Szymanski, B., Kuzmin, K., Xie, J., Lee, C., Blanche, T., Chaibub Neto, E., Huang, S.-C., Grabowski, T., et al. (2015). Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering. *Sci. Rep.* 5, 16361.
- Galatro, T.F., Holtman, I.R., Lerario, A.M., Vainchtein, I.D., Brouwer, N., Sola, P.R., Veras, M.M., Pereira, T.F., Leite, R.E.P., Möller, T., et al. (2017). Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat. Neurosci.* 20, 1162–1171.
- Gandal, M.J., Haney, J.R., Parikshak, N.N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A.J., Appadurai, V., Buil, A., Werge, T.M., et al.; CommonMind Consortium; PsychENCODE Consortium; iPSYCH-BROAD Working Group (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 359, 693–697.
- Gjoneska, E., Pfenning, A.R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.-H., and Kellis, M. (2015). Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* 518, 365–369.
- Götz, J., Bodea, L.-G., and Goedert, M. (2018). Rodent models for Alzheimer disease. *Nat. Rev. Neurosci.* 19, 583–598.
- Grubman, A., Chew, G., Ouyang, J.F., Sun, G., Choo, X.Y., McLean, C., Simmons, R.K., Buckberry, S., Vargas-Landin, D.B., Poppe, D., et al. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease

- p>reveals cell-type-specific gene expression regulation.
- Nat. Neurosci.*
- 22**
- , 2087–2097.
- Guadagno, N.A., Moriconi, C., Licursi, V., D'Acunto, E., Nisi, P.S., Carucci, N., De Jaco, A., Cacci, E., Negri, R., Lupo, G., and Miranda, E. (2017). Neuroserpin polymers cause oxidative stress in a neuronal model of the dementia FENIB. *Neurobiol. Dis.* **103**, 32–44.
- Guo, J.L., and Lee, V.M.Y. (2014). Cell-to-cell transmission of pathogenic proteins in neurodegenerative diseases. *Nat. Med.* **20**, 130–138.
- Haass, C., and Selkoe, D.J. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid  $\beta$ -peptide. *Nat. Rev. Mol. Cell Biol.* **8**, 101–112.
- Hargis, K.E., and Blalock, E.M. (2017). Transcriptional signatures of brain aging and Alzheimer's disease: What are our rodent models telling us? *Behav. Brain Res.* **322** (Pt B), 311–328.
- Haury, A.-C., Mordelet, F., Vera-Licona, P., and Vert, J.-P. (2012). TIGRESS: Trustful Inference of Gene Regulation Using Stability Selection. *BMC Syst. Biol.* **6**, 145.
- Holmes, A.P., Wong, S.Q., Pulix, M., Johnson, K., Horton, N.S., Thomas, P., de Magalhães, J.P., and Plagge, A. (2016). Reductions in hypothalamic Gfap expression, glial cells and  $\alpha$ -tanyocytes in lean and hypermetabolic Gnasxl-deficient mice. *Mol. Brain* **9**, 39.
- Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776.
- Ibrahim, F., Maragkakis, M., Alexiou, P., Maronski, M.A., Dichter, M.A., and Mourelatos, Z. (2013). Identification of in vivo, conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. *Cell Rep.* **3**, 301–308.
- Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.-L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* **345**, 455–459.
- Ishimura, R., Nagy, G., Dotu, I., Chuang, J.H., and Ackerman, S.L. (2016). Activation of GCN2 kinase by ribosome stalling links translation elongation with translation initiation. *eLife* **5**, e14295.
- Jack, C.R., Jr., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., et al.; Contributors (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* **14**, 535–562.
- Jankowsky, J.L., and Zheng, H. (2017). Practical considerations for choosing a mouse model of Alzheimer's disease. *Mol. Neurodegener.* **12**, 89.
- Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413.
- Jiao, S.-S., Bu, X.-L., Liu, Y.-H., Zhu, C., Wang, Q.-H., Shen, L.-L., Liu, C.-H., Wang, Y.-R., Yao, X.-Q., and Wang, Y.-J. (2016). Sex dimorphism profile of Alzheimer's disease-type pathologies in an APP/PS1 mouse model. *Neurotox. Res.* **29**, 256–266.
- Johnson, E.C.B., Dammer, E.B., Duong, D.M., Yin, L., Thambisetty, M., Troncoso, J.C., Lah, J.J., Levey, A.I., and Seyfried, N.T. (2018). Deep proteomic network analysis of Alzheimer's disease brain reveals alterations in RNA binding proteins and RNA splicing associated with disease. *Mol. Neurodegener.* **13**, 52.
- Johnson, E.C.B., Dammer, E.B., Duong, D.M., Ping, L., Zhou, M., Yin, L., Higinbotham, L.A., Guajardo, A., White, B., Troncoso, J.C., et al. (2019). A consensus proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *bioRxiv*. <https://doi.org/10.1101/802959>.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45** (D1), D353–D361.
- Kapasi, A., DeCarli, C., and Schneider, J.A. (2017). Impact of multiple pathologies on the threshold for clinically overt dementia. *Acta Neuropathol.* **134**, 171–186.
- Karch, C.M., Cruchaga, C., and Goate, A.M. (2014). Alzheimer's disease genetics: from the bench to the clinic. *Neuron* **83**, 11–26.
- Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522.
- Krämer, N., Schäfer, J., and Boulesteix, A.-L. (2009). Regularized estimation of large-scale gene association networks using graphical Gaussian models. *BMC Bioinformatics* **10**, 384.
- Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A., et al.; Alzheimer Disease Genetics Consortium (ADGC); European Alzheimer's Disease Initiative (EADI); Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE); Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES) (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430.
- Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Willighagen, E.L., Bohler, A., Mélius, J., Waagmeester, A., Sinha, S.R., Miller, R., et al. (2016). WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* **44** (D1), D488–D494.
- LaFerla, F.M., and Green, K.N. (2012). Animal models of Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**, a006320.
- Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H.-R., Jiang, J., Watt, A.T., Chun, S., Katz, M., et al. (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc. Natl. Acad. Sci. U S A* **110**, E4530–E4539.
- Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., and Zhang, K. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80.
- Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al.; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458.
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.
- Langfelder, P., Cantile, J.P., Chatzopoulou, D., Wang, N., Gao, F., Al-Ramahi, I., Lu, X.-H., Ramos, E.M., El-Zein, K., Zhao, Y., et al. (2016). Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat. Neurosci.* **19**, 623–633.
- Li, R., and Singh, M. (2014). Sex differences in cognitive impairment and Alzheimer's disease. *Front. Neuroendocrinol.* **35**, 385–403.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Logsdon, B.A., Gentles, A.J., Miller, C.P., Blau, C.A., Becker, P.S., and Lee, S.-I. (2015). Sparse expression bases in cancer reveal tumor drivers. *Nucleic Acids Res.* **43**, 1332–1344.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Mantel, N., and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**, 719–748.
- Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Kellis, M., Collins, J.J., and Stolovitzky, G.; DREAM5 Consortium (2012). Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804.

- Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 (Suppl 1), S7.
- Matarin, M., Salih, D.A., Yasvoina, M., Cummings, D.M., Guelfi, S., Liu, W., Nahaboo Solim, M.A., Moens, T.G., Paublete, R.M., Ali, S.S., et al. (2015). A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Rep.* 10, 633–644.
- Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 570, 332–337.
- Mayeux, R., and Stern, Y. (2012). Epidemiology of Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 2, a006239.
- Maze, I., Wenderski, W., Noh, K.-M., Bagot, R.C., Tzavaras, N., Purushothaman, I., Elsässer, S.J., Guo, Y., Ionete, C., Hurd, Y.L., et al. (2015). Critical role of histone turnover in neuronal transcription and plasticity. *Neuron* 87, 77–94.
- McKenzie, A.T., Moyon, S., Wang, M., Katsy, I., Song, W.-M., Zhou, X., Dammer, E.B., Duong, D.M., Aaker, J., Zhao, Y., et al. (2017). Multiscale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer's disease. *Mol. Neurodegener.* 12, 82.
- Meyer, P.E., Kontos, K., Lafitte, F., and Bontempi, G. (2007). Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J. Bioinform. Syst. Biol.* 2007, 79879.
- Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45 (D1), D183–D189.
- Mostafavi, S., Gaiteri, C., Sullivan, S.E., White, C.C., Tasaki, S., Xu, J., Taga, M., Klein, H.-U., Patrick, E., Komashko, V., et al. (2018). A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* 21, 811–819.
- Narayanan, M., Huynh, J.L., Wang, K., Yang, X., Yoo, S., McElwee, J., Zhang, B., Zhang, C., Lamb, J.R., Xie, T., et al. (2014). Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. *Mol. Syst. Biol.* 10, 743.
- Neuner, S.M., Heuer, S.E., Huentelman, M.J., O'Connell, K.M.S., and Kaczorowski, C.C. (2019). Harnessing genetic complexity to enhance translatability of Alzheimer's disease mouse models: a path toward precision medicine. *Neuron* 101, 399–411.e5.
- Nishimura, D. (2001). BioCarta. Biotech Softw. Internet Rep. 2, 117–120.
- Nixon, R.A. (2005). Endosome function and dysfunction in Alzheimer's disease and other neurodegenerative diseases. *Neurobiol. Aging* 26, 373–382.
- Nunn, N., Feetham, C.H., Martin, J., Barrett-Jolley, R., and Plagge, A. (2013). Elevated blood pressure, heart rate and body temperature in mice lacking the *XLas* protein of the *Gnas* locus is due to increased sympathetic tone. *Exp. Physiol.* 98, 1432–1445.
- Onos, K.D., Sukoff Rizzo, S.J., Howell, G.R., and Sasner, M. (2016). Toward more predictive genetic mouse models of Alzheimer's disease. *Brain Res. Bull.* 122, 1–11.
- Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandai, M.J., Hartl, C., Leppa, V., Ubieta, L.T., Huang, J., et al. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540, 423–427.
- Patrick, E., Olah, M., Taga, M., Klein, H.-U., Xu, J., White, C.C., Felsky, D., Gaiteri, C., Chibnik, L.B., Mostafavi, S., et al. (2017). A cortical immune network map identifies a subset of human microglia involved in Tau pathology. *bioRxiv*. <https://doi.org/10.1101/234351>.
- Pons, P., and Latapy, M. (2005). Computing communities in large networks using random walks. In *Computer and Information Sciences—ISCIS 2005*, I. Yolum, T. Güngör, F. Gürgeç, and C. Özturan, eds. (Berlin: Springer), pp. 284–293.
- Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I., et al. (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* 344, 519–523.
- Ramsden, M., Kotilinek, L., Forster, C., Paulson, J., McGowan, E., SantaCruz, K., Guimaraes, A., Yue, M., Lewis, J., Carlson, G., et al. (2005). Age-dependent neurofibrillary tangle formation, neuron loss, and memory impairment in a mouse model of human tauopathy (P301L). *J. Neurosci.* 25, 10637–10647.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Rojo, A.I., Pajares, M., Rada, P., Nuñez, A., Nevado-Holgado, A.J., Killik, R., Van Leuven, F., Ribe, E., Lovestone, S., Yamamoto, M., and Cuadrado, A. (2017). NRF2 deficiency replicates transcriptomic changes in Alzheimer's patients and worsens APP and TAU pathology. *Redox Biol.* 13, 444–451.
- Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. *Nat. Med.* 10 (Suppl), S10–S17.
- Rosvall, M., and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U S A* 105, 1118–1123.
- Rothman, S.M., Tanis, K.Q., Gandhi, P., Malkov, V., Marcus, J., Pearson, M., Stevens, R., Gilliland, J., Ware, C., Mahadomrongkul, V., et al. (2018). Human Alzheimer's disease gene expression signatures and immune profile in APP mouse models: a discrete transcriptomic view of A $\beta$  plaque pathology. *J. Neuroinflammation* 15, 256.
- Roussel, B.D., Kruppa, A.J., Miranda, E., Crowther, D.C., Lomas, D.A., and Marciniak, S.J. (2013). Endoplasmic reticulum dysfunction in neurological disease. *Lancet Neurol.* 12, 105–118.
- Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Winblad, B., and Saido, T.C. (2010). GeneCards Version 3: the human gene integrator. *Database* 2010, baq020.
- Sala Frigerio, C., Wolfs, L., Fattorelli, N., Thrupp, N., Voytyuk, I., Schmidt, I., Mancuso, R., Chen, W.-T., Woodbury, M.E., Srivastava, G., et al. (2019). The major risk factors for Alzheimer's disease: age, sex, and genes modulate the microglia response to A $\beta$  plaques. *Cell Rep.* 27, 1293–1306.e6.
- Sasaguri, H., Nilsson, P., Hashimoto, S., Nagata, K., Saito, T., De Strooper, B., Hardy, J., Vassar, R., Winblad, B., and Saido, T.C. (2017). APP mouse models for Alzheimer's disease preclinical studies. *EMBO J.* 36, 2473–2487.
- Scheltens, P., Blennow, K., Breteler, M.M.B., de Strooper, B., Frisoni, G.B., Salloway, S., and Van der Flier, W.M. (2016). Alzheimer's disease. *Lancet* 388, 505–517.
- Seyfried, N.T., Dammer, E.B., Swarup, V., Nandakumar, D., Duong, D.M., Yin, L., Deng, Q., Nguyen, T., Hales, C.M., Wingo, T., et al. (2017). A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. *Cell Syst.* 4, 60–72.e4.
- Skene, N.G., and Grant, S.G.N. (2016). Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.* 10, 16.
- Song, W.-M., and Zhang, B. (2015). Multiscale embedded gene co-expression network analysis. *PLoS Comput. Biol.* 11, e1004574.
- Srinivasan, K., Friedman, B.A., Larson, J.L., Lauffer, B.E., Goldstein, L.D., Appling, L.L., Borneo, J., Poon, C., Ho, T., Cai, F., et al. (2016). Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. *Nat. Commun.* 7, 11295.
- Stephenson, J., Nutma, E., van der Valk, P., and Amor, S. (2018). Inflammation in CNS neurodegenerative diseases. *Immunology* 154, 204–219.
- Swartzlander, D.B., Propson, N.E., Roy, E.R., Saito, T., Saido, T., Wang, B., and Zheng, H. (2018). Concurrent cell type-specific isolation and profiling of mouse brains in inflammation and Alzheimer's disease. *JCI Insight* 3, e121109.
- Traag, V.A., and Bruggeman, J. (2009). Community detection in networks with positive and negative links. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 80, 036115.



- Tryka, K.A., Hao, L., Sturcke, A., Jin, Y., Wang, Z.Y., Ziyabari, L., Lee, M., Popova, N., Sharopova, N., Kimura, M., and Feolo, M. (2014). NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res.* 42, D975–D979.
- Ung, D.C., Iacono, G., Méziane, H., Blanchard, E., Papon, M.-A., Selten, M., van Rhijn, J.-R., Montjean, R., Rucci, J., Martin, S., et al. (2018). Ptchd1 deficiency induces excitatory synaptic and cognitive dysfunctions in mouse. *Mol. Psychiatry* 23, 1356–1367.
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Vied, C., Ray, S., Badger, C.-D., Bundy, J.L., Arbeitman, M.N., and Nowakowski, R.S. (2016). Transcriptomic analysis of the hippocampus from six inbred strains of mice suggests a basis for sex-specific susceptibility and severity of neurological disorders. *J. Comp. Neurol.* 524, 2696–2710.
- Wang, M., Roussos, P., McKenzie, A., Zhou, X., Kajiwara, Y., Brennand, K.J., De Luca, G.C., Crary, J.F., Casaccia, P., Buxbaum, J.D., et al. (2016). Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med.* 8, 104.
- Wang, M., Beckmann, N.D., Roussos, P., Wang, E., Zhou, X., Wang, Q., Ming, C., Neff, R., Ma, W., Fullard, J.F., et al. (2018). The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci. Data* 5, 180185.
- Wang, M., Li, A., Sekiya, M., Beckmann, N.D., Quan, X., Schrode, N., Fernando, M.B., Yu, A., Zhu, L., Cao, J., et al. (2019). Molecular networks and key regulators of the dysregulated neuronal system in Alzheimer's disease. *bioRxiv*. <https://doi.org/10.1101/788323>.
- Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573.
- Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezchnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 153, 707–720.
- Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34, 11929–11947.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
AMP-AD individual AD coexpression modules	This study	<a href="https://doi.org/10.7303/syn10309369.1">https://doi.org/10.7303/syn10309369.1</a>
AMP-AD aggregate AD coexpression modules	This study	<a href="https://doi.org/10.7303/syn11932957.1">https://doi.org/10.7303/syn11932957.1</a>
AMP-AD aggregate AD coexpression modules gene-set enrichment results	This study	<a href="https://doi.org/10.7303/syn11954640.1">https://doi.org/10.7303/syn11954640.1</a>
Differential expression meta-analysis of reprocessed RNASeq data from AMP-AD (all 7 brain regions)	This study	<a href="https://doi.org/10.7303/syn11914606">https://doi.org/10.7303/syn11914606</a>
The Religious Order Study and the Memory and Aging Project (ROSMAP) RNA-seq	De Jager et al., 2018	<a href="https://doi.org/10.7303/syn3388564">https://doi.org/10.7303/syn3388564</a>
The Mount Sinai Brain Bank (MSBB) RNA-seq	Wang et al., 2018	<a href="https://doi.org/10.7303/syn3157743">https://doi.org/10.7303/syn3157743</a>
The Mayo clinic RNA-seq	Allen et al., 2016	<a href="https://doi.org/10.7303/syn5550404">https://doi.org/10.7303/syn5550404</a>
Mouse orthologs for human genes extracted using the HCOP tool	This study	<a href="https://doi.org/10.7303/syn17010253.1">https://doi.org/10.7303/syn17010253.1</a>
The AD Cross Species study (mouse model RNA-seq reprocessing, differential expression analysis, and mouse-human overlaps).	This study	<a href="https://doi.org/10.7303/syn16779040">https://doi.org/10.7303/syn16779040</a>
Software and Algorithms		
R language v3.3.0-3.5.1	The R Foundation for Statistical Computing	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
Picard v2.18.2-2.2.4	Broad Institute	<a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a>
STAR v2.5.1b	Dobin et al., 2013 (PMID: 23104886)	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
voom-limma	Ritchie et al., 2015 (PMID: 25605792)	<a href="https://bioconductor.org/packages/release/bioc/html/limma.html">https://bioconductor.org/packages/release/bioc/html/limma.html</a>
rmeta	CRAN	<a href="https://cran.r-project.org/web/packages/rmeta/index.html">https://cran.r-project.org/web/packages/rmeta/index.html</a>
MEGENA	Song and Zhang, 2015 (PMID: 26618778)	<a href="https://cran.r-project.org/web/packages/MEGENA/index.html">https://cran.r-project.org/web/packages/MEGENA/index.html</a>
WINA	Wang et al., 2016 (PMID: 27799057)	<a href="https://doi.org/10.7303/syn7221264.2">https://doi.org/10.7303/syn7221264.2</a>
metanetwork	This study	<a href="https://github.com/Sage-Bionetworks/metanetwork">https://github.com/Sage-Bionetworks/metanetwork</a>
rWGCNA	Parikshak et al., 2016 (PMID: 27919067)	<a href="https://github.com/dhglab/Genome-wide-changes-in-lncRNA-alternative-splicing-and-cortical-patterning-in-autism">https://github.com/dhglab/Genome-wide-changes-in-lncRNA-alternative-splicing-and-cortical-patterning-in-autism</a>
speakEasy	Gaiteri et al., 2015 (PMID: 26549511)	<a href="http://www.cs.rpi.edu/~szymansk/SpeakEasy">http://www.cs.rpi.edu/~szymansk/SpeakEasy</a>
igraph	Csardi and Nepusz, 2006. <a href="https://igraph.org">https://igraph.org</a>	<a href="https://cran.r-project.org/web/packages/igraph/index.html">https://cran.r-project.org/web/packages/igraph/index.html</a>
Snakemake v4.8.0	Köster and Rahmann, 2012 (PMID: 22908215)	<a href="https://snakemake.readthedocs.io/en/stable">https://snakemake.readthedocs.io/en/stable</a>
GEOquery R package v2.42.0	Davis and Meltzer, 2007 (PMID: 17496320)	<a href="https://bioconductor.org/packages/release/bioc/html/GEOquery.html">https://bioconductor.org/packages/release/bioc/html/GEOquery.html</a>
NCBI SRA toolkit v2.8.2.1	N/A	<a href="https://github.com/ncbi/sra-tools">https://github.com/ncbi/sra-tools</a>
samtools v0.1.5	Li et al., 2009 (PMID: 19505943)	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
HTSeq v0.6.0	Anders et al., 2015 (PMID: 25260700)	<a href="https://htseq.readthedocs.io/en/master">https://htseq.readthedocs.io/en/master</a>
DESeq2 v1.18.1	Love et al., 2014 PMID: 25516281	<a href="http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html">http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html</a>

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Python v.2.7.12	N/A	<a href="https://www.python.org/">https://www.python.org/</a>
HGNC Comparison of Orthology Predictions (HCOP)	N/A	<a href="https://www.genenames.org/tools/hcop">https://www.genenames.org/tools/hcop</a>

**RESOURCE AVAILABILITY**

**Lead Contact**

Further information and requests for data or other resources should be directed to and will be fulfilled by the Lead Contact, Benjamin A. Logsdon ([ben.logsdon@sagebionetworks.org](mailto:ben.logsdon@sagebionetworks.org)).

**Materials Availability**

This study did not generate new unique reagents.

**Data and code availability**

All study data is available on the AMP-AD Knowledge Portal (<https://doi.org/10.7303/syn2580853>). The complete list of deposited data (with links), software, and relevant algorithms is provided in the Key Resources Table. Specifically, all original and re-processed RNA-seq data for the ROSMAP, MSBB, and Mayo clinic cohorts are available, along with the results of our differential expression meta-analysis across the three cohorts and 7 brain regions. Comprehensive results for coexpression analysis (2978 modules) generated using 5 algorithms across 7 brain regions in the three cohorts are also provided, along with the resulting aggregate AD coexpression modules (30 modules). Lastly, we also have deposited all data from the reprocessing of mouse model RNA-seq, along with the associated differential expression analysis, and complete results of mouse-human overlaps. An R package with all code for the metanetwork algorithm is available at <https://github.com/Sage-Bionetworks/metanetwork>, and a toolkit for integrating meta-network with AWS high performance compute cluster cfnccluster, and Synapse is also made available (<https://github.com/Sage-Bionetworks/metanetworkSynapse>). All code used to generate aggregate modules and figures are available in this R package: <https://github.com/Sage-Bionetworks/AMPAD>, with the following notebook collating the primary results: <https://github.com/Sage-Bionetworks/AMPAD/blob/master/manuscript.Rmd>.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

**Human subject data**

Details of sample collection, postmortem sample descriptions, tissue and RNA preparation, library preparation and sequencing, and sample QC are provided in previously published work (Allen et al., 2016; De Jager et al., 2018; Wang et al., 2018). Overall, the combined cohort consisted of 2,114 total samples from 1,234 subjects (Table 1; Table S1), including 478 AD cases, 300 controls (without AD pathology), and 456 subjects with progressive supranuclear palsy or other neurodegenerative pathologies. Sub-samples were selected to harmonize the late onset AD (LOAD) case - control definition across the three studies for all differential expression analyses. To compare analysis results across studies and to get an understanding of LOAD biology across different tissues, we harmonized the LOAD definition across three studies. The motivation was to define LOAD cases as those with both clinical and neuropathological evidence for definitive late onset Alzheimer's disease - i.e., a high burden of neurofibrillary tangles, neuritic amyloid plaques, and cognitive impairment with little evidence of other pathology (Jack et al., 2018). Controls were concordantly defined as patients with a low burden of plaques and tangles, as well as no evidence of cognitive impairment if available. As such, for the ROSMAP study, we had individuals with a Braak neurofibrillary tangle (NFT) score (Braak et al., 2006) greater than or equal to 4, CERAD score less than or equal to 2, and a cognitive diagnosis of probable AD with no other causes as LOAD cases, Braak less than or equal to 3, CERAD score greater than or equal to 3, and cognitive diagnosis of 'no cognitive impairment' as LOAD controls. For MSBB, we analogously defined LOAD cases as those with CDR score greater than or equal to 1, Braak score greater than or equal to 4, and CERAD neuritic and cortical plaque score greater than or equal to 2 as LOAD cases, and CDR scores less than or equal to 0.5, Braak less than or equal to 3, and CERAD less than or equal to 1 as LOAD controls. It is to note here that the definitions of CERAD differs between ROSMAP and MSBB studies. For the Mayo Clinic RNASeq study, cases were defined based on neuropathology, with LOAD cases being based on Braak score greater than or equal to 4 and CERAD neuritic and cortical plaque score greater than 1 whereas LOAD controls being those defined as Braak less than or equal to 3, and CERAD less than 2. Further details concerning the diagnosis in the Mayo RNASeq study have been previously published (Allen et al., 2018). Summary of sample sizes of AD cases and controls are shown in Table 1.

**Mouse Experimental Model Data**

Figure 2 depicts the overall analysis pipeline for collecting and processing mouse studies, and examining overlaps with human coexpression modules. A total of 96 studies, encompassing data from 2279 mouse tissue samples, were analyzed (Table S3). These data

were collected from three sources: the Gene Expression Omnibus (GEO) database (83 studies), the AMP-AD Knowledge portal (syn5550383) (6 studies), and through personal communication (7 studies). We searched the GEO database on September 12, 2017 using the keywords “brain,” “mouse,” and “expression profiling by high throughput sequencing,” identifying 881 studies for initial consideration. All studies were next indexed using high frequency terms, and secondary filtering was based on manually curated keywords relevant to AD pathophysiology (Karch et al., 2014) (Table S3). Lastly, the filtered list of 349 studies was reviewed by the study team. We excluded studies, or in some cases specific samples, involving (1) tissue source other than nervous system, (2) organisms other than mouse, or (3) non-coding RNA. From the remainder, 79 GEO studies were selected for inclusion based on relevance to AD, neurodegeneration, or related mechanisms. All included studies had publicly available RNA-seq data derived from either mouse brain or brain-derived cell lines. Following these initial searches, we discovered only a single eligible RNA-seq study for analysis of aging-associated expression changes (GSE61915). Given the importance of aging in AD, we identified and included an additional 5 expression array profiling studies related to brain aging. Table S4 details all studies included in this analysis, including data source, citations, and relevant keywords.

## METHOD DETAILS

As our study exclusively involves the analysis of previously generated data, all relevant further method details are described below.

### Human RNA-Seq Reprocessing, library normalization and covariates adjustment

To avoid some of the technical variabilities arising due to RNA-seq alignment and quantification, and also to account for some of the technical variabilities we reprocessed and realigned all the RNA-Seq reads from the source studies (Allen et al., 2016; De Jager et al., 2018; Wang et al., 2018). The reprocessing was done using a consensus set of tools with only library type-specific parameters varying between pipelines. Picard (<https://broadinstitute.github.io/picard/>) was used to generate FASTQs from source BAMs. Generated FASTQ reads were aligned to the GENCODE24 (GRCh38) reference genome using STAR (Dobin et al., 2013) and gene counts were computed for each sample. To evaluate the quality of individual samples and to identify potentially important covariates for expression modeling, we computed two sets of metrics using the CollectAlignmentSummaryMetrics and CollectRnaSeqMetrics functions in Picard. To account for differences between samples, studies, experimental batch effects and unwanted RNA-Seq specific technical variations we performed library normalization and covariate adjustments for each study separately using fixed/mixed effects modeling. The workflow consists of the following steps: (i) gene filtering: Genes that are expressed more than 1 CPM (read Counts Per Million total reads) in at least 50% of samples in each tissue and diagnosis category was used for further analysis, (ii) conditional quantile normalization was applied to account for variations in gene length and GC content, (iii) sample outlier detection using principal component analysis and clustering, (iv) Covariates identification and adjustment, where confidence of sampling abundance were estimated using a weighted linear model using voom-limma package in bioconductor (Ritchie et al., 2015). For most analyses, we perform a variant of fixed/mixed effect linear regression as shown here: gene expression  $\sim$  Diagnosis + Sex + covariates + (1|Donor) or gene expression  $\sim$  Diagnosis x Sex + covariates + (1|Donor), where each gene is linearly regressed independently with Diagnosis, a variable explaining the AD status of an individual, identified covariates, and donor information as a random effect. Observation weights (if any) were calculated using the voom-limma (Ritchie et al., 2015) pipeline such that observations with higher presumed precision will be up-weighted in the linear model fitting process. All workflows were applied separately for each of the three studies.

### Network Inference and Module Identification

We apply 5 distinct network module identification methodologies to each of the 7 tissue specific expression datasets. This includes MEGENA (Song and Zhang, 2015), WINA (Wang et al., 2016), metanetwork, rWGCNA (Parikshak et al., 2016), speakEasy (Gaiteri et al., 2015), and a novel “metanetwork” algorithm to characterize a comprehensive landscape of transcriptomic variation across the seven brain regions and three studies. Briefly, MEGENA (Song and Zhang, 2015) is a method that infers a sparse graph based on a distance to define multiscale module definitions from coexpression data. Speakeasy is a label propagation method to identify robust coexpression modules that are identified both top up and bottom down (Gaiteri et al., 2015), rWGCNA is a version of WGCNA (Langfelder and Horvath, 2008) that includes bootstrapping to identify robust modules, WINA is also a variation on WGCNA that includes a modified tree cutting method to identify modules (Wang et al., 2016).

The metanetwork inference methodology is inspired by the DREAM5 method (Marbach et al., 2012), where ensemble inference methodologies were identified as more robust for identification of gene-gene interactions from coexpression data. We constructed a statistical network of gene co-expression using an ensemble network inference algorithm. Briefly, we apply nine distinct gene co-expression network inference methodologies ARACNE (Margolin et al., 2006a), Genie3 (Huynh-Thu et al., 2010), Tigress (Haury et al., 2012), Sparrow (Logsdon et al., 2015), Lasso (Krämer et al., 2009), Ridge (Krämer et al., 2009), mnet (Meyer et al., 2007), c3net (Altay and Emmert-Streib, 2010) and WGCNA (Langfelder and Horvath, 2008) and rank the edge lists from each method based on the method specific edge weights, identify a mean rank for each edge across methods, then identify the total number of edges supported by the data with Bayesian Information Criterion for local neighborhood selection with linear regression. We identify metanetwork modules in each tissue type based on the inferred network topology with a consensus clustering algorithm (Wilkerson and Hayes, 2010) applied to multiple individual module identification methods. We ran individual network clustering methods applied to each of the seven network topologies. These methods included CFinder (Adamcsek et al., 2006), GANXiS (Gaiteri et al., 2015), a fast

greedy algorithm (Clauset et al., 2004), InfoMap (Rosvall and Bergstrom, 2008), LinkCommunities (Ahn et al., 2010), Louvain (Blondel et al., 2008), Spinglass (Traag and Bruggeman, 2009), and Walktrap (Pons and Latapy, 2005), methods. All implementations are from the igraph package (Csardi et al.) in R.

### Aggregate human coexpression module identification

For all 2978 modules identified across tissues, we first identify which modules are enriched for  $> 1$  AD specific differential expressed gene set from the DEG meta-analysis (see below). This restricts the total number of individual modules to 660 that show evidence of differential expression as a function of disease status. Next, we construct a within tissue module graph using a Fisher's exact test for pairwise overlap of gene sets between each pair of these 660 individual modules. We then apply the edge betweenness graph clustering method (Pons and Latapy, 2005) to identify aggregate modules from these module graphs that represent meta modules that are both differentially expressed and identified by multiple independent module identification algorithms. With this approach we identify 30 aggregate module definitions across the seven tissue types and three studies (see Data and code availability).

In order to test the cross-region and cross-study robustness of these AD associated consensus clusters, we built aggregate modules from all of the individual modules generated for each brain region then compared these modules to the AD associated aggregate modules (Figure S2A). We find that there is significant overlap between the AD associated aggregate modules (AD) and the modules constructed with all individual modules (All) into the five consensus clusters. Those modules non-associated with AD do not show similar conservation across brain region, increasing confidence that the AD-associated modules are robust to technical and study artifacts.

### Mouse RNA-seq re-processing

A unified RNA-seq analysis pipeline was used for reprocessing of all datasets, with the exception of 2 HD studies where count files were downloaded directly from GEO. Data processing leveraged the cloud formation cluster at Amazon Web Services. We first created one EC2 master instance (m3.xlarge) which was used to launch hundreds of EC2 computing nodes (c3.8xlarge). Next, each computing node was assigned to process one sample using our customized RNA-seq pipeline, as implemented using Snake-make v.4.8.0 (Köster and Rahmann, 2012). For samples in AMP-AD studies, the pipeline begins with downloading BAM files from the AMP-AD Knowledge Portal using the Synapse python client. The BAM files are then converted to fastq files using Picard SamToFastq command v.2.18.2 (<http://broadinstitute.github.io/picard>). For GEO studies, SRA files were downloaded from the database using the GEOquery R package v.2.42.0 (Davis and Meltzer, 2007), and fastq files were generated using the fastq-dump command from the NCBI SRA toolkit v2.8.2.1 (Andrews, 2010). Alignment to the mouse reference genome GRCm38 (mm10) was implemented using STAR v.2.5.1b (Dobin et al., 2013), and BAM file reads were subsequently sorted by coordinate using samtools v0.1.5 (Li et al., 2009). Genes were quantified using either HTSeq v0.6.0 (Anders et al., 2015) or using the 'quantMode' option from the STAR aligner which utilizes HTSeq algorithm and produces similar results. Results were uploaded to the Synapse portal using the python client.

### Nomenclature for annotation of mouse differentially expressed gene sets

To facilitate ease of use and repurposing of the data, each mouse DEG set was assigned a unique identifier (M###), and we also developed a descriptive nomenclature for annotation. Each gene set received a label taking the form: *category\_experimental condition\_sex\_age\_brain region\_cell type\_transgene*. In this standardized annotation, "category" denotes the relevant neurologic disease (e.g., AD, HD, SCA, ALS) or "other" for gene manipulations not directly linked to human disease, along with the specific "experimental condition" describing the mouse genotype or treatment condition. We also note "sex" (M or F), "age" (months), and where applicable, "brain region" (e.g., hippocampus), "cell type" (e.g., neuron, microglia). In the case of AD mouse models, we also annotate "transgene," to differentiate "APP," "Tau" (MAPT), or "other" models. If unknown or not applicable, the relevant field(s) are replaced with "na." These annotations and conventions are used throughout our supplementary tables and files. The estimated pathologic burden (plaques/tangles and neuronal loss) in APP and MAPT models was annotated based on Alzforum (<https://www.alzforum.org/research-models/>) (Table S4).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Meta-Differential expression analysis on human expression profiles

All the differential and meta-differential expression analysis were performed as weighted fixed/mixed effect linear models using the voom-limma (Ritchie et al., 2015) package in R. For each gene, linear regression was fit with biological and technical covariates that were associated with the top principal components of the expression data, as identified above. Two of the three studies - MSBB and Mayo RNaseq - obtained more than one tissue from the same donors. Therefore, except the ROSMAP study, donor-specific effects were explicitly modeled as random effects. Different models were built for understanding the effects of diagnosis and sex-specific diagnosis effects. Depending on the model, coefficients related to either diagnosis or diagnosis times sex was statistically tested for being non-zero, implying an estimated effect for the primary variable of interest is above and beyond any other effect from the covariates. This test produces a t-statistic (then moderated in a Bayesian fashion) and corresponding p value. P values were then adjusted for multiple hypothesis testing using false discovery rate (FDR) estimation, and the differentially expressed genes were determined as those with an estimated FDR below, or at, 5% with a corresponding absolute expression and fold-change cutoffs. To identify genes with evidence for change in expression across studies, we next performed a meta-analysis using a random effect

and fixed effect models using *rmeta* R package (<https://cran.r-project.org/web/packages/rmeta/index.html>). The random effect model was selected as a conservative approach to correct for variation across studies. Sample size and characteristics of data included in the differential expression analyses are detailed in Table 1.

### Human coexpression module enrichment analysis

Aggregate modules were interpreted using functional and cell type enrichment analysis. We performed a battery of enrichment tests to understand biological functionality, including evaluating primary hypotheses previously implicated by genetic findings in AD research, performing exploratory analyses of a large number of gene sets, and performing enrichment for brain tissue specific cell types. Among gene sets, we considered curated AD genes from GeneCards (Safran et al., 2010), Panther (Mi et al., 2017), dbGaP (Tryka et al., 2014), IGAP (Lambert et al., 2013), OMIM (Amberger et al., 2015), Biocarta (Nishimura, 2001), Wikipathways (Kutmon et al., 2016), and KEGG (Kanehisa et al., 2017). We also examined cell type markers from a mouse bulk brain tissue RNaseq atlas (Zhang et al., 2014) and human single-cell RNaseq (Lake et al., 2018). AMP-AD specific gene sets were constructed by taking the union of gene set definitions reported in each of the following reports: RNA-binding protein modules (Johnson et al., 2018), oligodendroglial modules from MSSM (McKenzie et al., 2017), AD versus Control oligodendroglial modules in the Mayo RNaseq study (Allen et al., 2018), and Module 109 from the ROSMAP study (Mostafavi et al., 2018). Genes not measured in our data are filtered from the annotated gene sets. Annotated gene sets with less than 10% of genes expressed in our datasets were removed. Fisher's exact test was used to test enrichment of each gene set with the annotated set. Resulting p values were corrected independently for each set using Benjamini-Hochberg method for significance testing, owing to the differences in their hypothesis. Gene sets that had a minimum overlap of at least 3 genes were considered for further interpretation.

### Differential expression analysis of mouse RNA-seq/microarray

Differential gene expression analysis was conducted using DESeq2 v1.18.1 (Love et al., 2014). For the limited number of microarray studies, pre-processed intensities available from the series matrix files were downloaded from GEO and normalized using quantile normalization, followed by differential expression analysis using the *limma* package v3.4.2 (Ritchie et al., 2015). For each study (Tables S3 and S4), experimental and control pairs were manually curated. We required a minimum of  $n = 2$  samples for each group (experimental and control); the average for all samples included in each comparison was  $n = 8.4$  (range = 4–28 total samples). Overall, 376 mouse experimental comparisons were curated for computation of differentially expressed gene sets (DEGs), applying a false-discovery rate (FDR) threshold of 1% and minimum fold-change of 1.2. The t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm (van der Maaten and Hinton, 2008) was applied to DEGs from all studies (logarithm-transformed fold-change), using the *Rtsne* function in R. We excluded all DEG sets from consideration consisting of fewer than 10 conserved mouse genes, resulting in 251 sets of DEGs for consideration in our subsequent analyses. All analysis was done using R v3.4.2, and Python v2.7.12. Table S4 details all DEG sets meeting these criteria, including data sources, and also enumerates the sample size ( $n$ ) for all comparisons.

### Statistical analysis of mouse-human overlaps

Mouse orthologs for all human genes were extracted using the HCOP tool available from the HUGO Gene Nomenclature Committee (see Data and code availability). Due to wide diversity of potential AD mechanisms, along with the number and heterogeneity of mouse models included in our analyses, it is not possible to confidently select the specific human gene-ortholog pairs in most cases where human genes map to multiple mouse orthologs. Therefore, in our primary analysis, we considered all possible human-mouse ortholog pairs (Table S5 and all results reported in manuscript). In a secondary analysis (Table S6), we performed a sensitivity test by restricting to human-mouse gene ortholog pairs with 1:1 relationships. In cases with multiple homologs, we selected the highest-confidence mouse ortholog for each human gene based on HCOP. Using the hypergeometric test, we determined the significance of overlap between each of 251 mouse DEG sets (above) and the 30 human gene coexpression modules (mouse orthologs) using the *phyper* function in R. The Benjamini-Hochberg method was applied to adjust for multiple comparisons, using the *p.adjust* function. All p values reported in the text were adjusted in this manner. Overall, out of 1569 significant module-DEG set overlaps ( $p_{adj} < 0.01$ ) detected in our primary analysis, 1306 (83%) are recovered by the more stringent sensitivity test using 1:1 human-mouse ortholog mapping. Consistency is even higher among 1058 highly significant overlaps ( $p_{adj} < 0.00001$ ) from our primary analysis, for which 1014 (95.8%) are recapitulated using the more stringent 1:1 ortholog definition. Overlap significance was visualized using heatmaps, implemented with *pheatmap* function in R, using Manhattan distance (for both rows and columns) and Ward clustering. In order to determine whether mouse-human overlapping genes also shared expression changes in the same direction, we computed the concordance score for each overlap. Specifically, the concordance score is the percentage of genes in the concordant direction weighted by the significance based on the hypergeometric test, which is computed as follows:

$$WeightedConc = -\log_{10}(p_{adj}) \times (ConcUp + ConcDown),$$

where

$$ConcUp = \frac{\# \text{ overlapped genes up in both mouse and human}}{\# \text{ overlapped genes up in mouse}},$$

$$\text{ConcDown} = \frac{\# \text{ overlapped genes down in both mouse and human}}{\# \text{ overlapped genes down in mouse}}$$

The overlap was considered concordant when the weighted concordance is greater (or less) than half standard deviation from the median, and  $p_{\text{adj}} \leq 0.01$  for the up- or down- differential expression. All computation and calculations were carried out in the R language for statistical computing (version 3.3.0 - 3.5.1) and Python v.2.7.12.